

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 1, 2002, 20:06:13 ; Search time 1485.96 Seconds
(without alignments)
8404.240 Million cell updates/sec

Title: US-09-586-106-62

Perfect score: 757

Sequence: 1 gaaagaggtlgtgaagctc.....atcaagattcacaaggt 757

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 1472140 segs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenBank.*
1: gb_da.*
2: gb_hg.*
3: gb_in.*
4: gb_ov.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sv.*
13: gb_un.*
14: gb_vl.*
15: em_da.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_com.*
20: em_or.*
21: em_ov.*
22: em_pat.*
23: em_ph.*
24: em_pl.*
25: em_ro.*
26: em_sts.*
27: em_sy.*
28: em_un.*
29: em_vl.*
30: em_htgo_hum.*
31: em_htgo_inv.*
32: em_htgo_rod.*
33: em_htg_hum.*
34: em_htg_inv.*
35: em_htg_rod.*
36: em_htg_other.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	571.4	75.5	65979	AF254799	AF254799 Hordeum v
2	389.4	51.4	83646	AB005248	AB005248 Arabidops
3	381.6	50.4	49981	AX059550	AX059550 Sequence
4	381.6	50.4	54708	AC007209	AC007209 Arabidops
5	381.6	50.4	63292	F14616	AF147260 Arabidops
6	381.6	50.4	198176	ATCHR1V17	AL161505 Arabidops
7	380.8	50.2	95685	AC007187	AC007187 Arabidops
8	380.8	50.2	40480	AX059456	AX059456 Sequence
9	380	50.2	76911	AB016828	AC016828 Arabidops
10	380	50.2	85962	AB046431	AB046431 Arabidops
11	380	50.2	87814	AB046426	AB046426 Arabidops
12	380	50.2	95463	F1809	AF1809 Arabidops
13	380	50.2	118986	AC007534	AC007534 Arabidops
14	378.4	50.0	54573	T619	AB026642 Arabidops
15	378.4	50.0	77287	AB026642	AB026642 Arabidops
16	378.4	50.0	183181	ATCHR1V19	AL161507 Arabidops
17	376.8	49.8	18813	T9E19	AF147265 Arabidops
18	376.8	49.8	35313	AB028613	AB028613 Arabidops
19	376.8	49.8	51248	AB046437	AB046437 Arabidops
20	376.8	49.8	67701	AC007155	AC007155 Arabidops
21	376.8	49.8	77204	AB046438	AB046438 Arabidops
22	376.8	49.8	108227	ATF7P3	AL138663 Arabidops
23	376.8	49.8	121501	AC069557	AC069557 Genomic S
24	375.2	49.6	37010	AX059491	AX059491 Sequence
25	375.2	49.6	38519	AX059490	AX059490 Sequence
26	375.2	49.6	42112	AX059547	AX059547 Sequence
27	375.2	49.6	47347	AX059509	AX059509 Sequence
28	375.2	49.6	48422	AX059509	AX059509 Sequence
29	375.2	49.6	79122	ATAC09261	AC09261 Arabidops
30	375.2	49.6	88126	F3D18	AF147261 Arabidops
31	375.2	49.6	94905	F2112	AF147261 Arabidops
32	375.2	49.6	94905	AC019013	AL161504 Arabidops
33	375.2	49.6	114056	ATCHR1V16	AL161504 Arabidops
34	375.2	49.6	196766	ATCHR1V16	AL161504 Arabidops
35	375.2	49.6	196766	ATCHR1V16	AL161504 Arabidops
36	375.2	49.6	196766	ATCHR1V16	AL161504 Arabidops
37	372	49.1	107487	AC087569	AC087569 Arabidops
38	371.2	49.0	39104	AX059494	AX059494 Sequence
39	371.2	49.0	87219	AP002054	AP002054 Arabidops
40	371.2	49.0	129021	ATT32A11	AL138653 Arabidops
41	370.4	48.9	36096	AX059471	AX059471 Sequence
42	370.4	48.9	38243	AX059470	AX059470 Sequence
43	370.4	48.9	103931	AC020646	AC020646 Genomic S
44	370.4	48.9	108720	AB046436	AB046436 Arabidops
45	370.4	48.9	134499	AC007505	AC007505 Arabidops

ALIGNMENTS

RESULT 1
AF254799/c
LOCUS AF254799
DEFINITION Hordeum vulgare (onoplas) Intrinsic protein 1 (TIP1), complete cds.
ACCESSION AF254799
VERSION AF254799.1 GI:9623334
KEYWORDS
SOURCE
ORGANISM Hordeum vulgare
Barley.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Hordeum.
REFERENCE
AUTHORS Shirasu K., Schulman A.H., Lahaye T. and Schulze-Lefert P.
TITLE A contiguous 66-kb barley DNA sequence provides evidence for
reversible genome expansion
JOURNAL Genome Res. 10 (7), 908-915 (2000)
MEDLINE 20359834
REFERENCE 2 (bases 1 to 65979)

AUTHORS Shirasu, K., Schulman, A.H., Lahaye, T. and Schulze-Lefert, P.
TITLE Direct Submission
JOURNAL Submitted (11-APR-2000) The Sainsbury Laboratory, John Innes
Centre, Colney Lane, Norwich NR4 7UH, UK

FEATURES
Source Location/Qualifiers
1. 65979

misc_feature
/organism="Hordeum vulgare"
/db_xref="taxon:4513"
/chromosome="2"
/map="2HL: MNG87"
402..5501
/note="similar to BAGY-2 polyprotein"

LTR
5608..6244
/note="BAGY-2 LTR-1"

LTR
6245..11204
/note="Sukula LTR-1"

LTR
11205..11757
/note="BAGY-2 LTR-1"

LTR
11758..13280
/note="BAGY-2 LTR-2"

LTR
13281..13627
/note="BAGY-2 LTR-1"

LTR
14378..14966
/note="Sabrina LTR-1"

LTR
14967..16796
/note="BARE-1 LTR-1"

LTR
16797..17590
/note="Sabrina LTR-1"

LTR
19160..20144
/note="Sabrina LTR-2"

LTR
20145..24048
/note="Sukula LTR-2"

LTR
24049..23865
/note="BARE-1 LTR-2"

LTR
25866..27570
/note="Sukula LTR-2"

LTR
27571..28211
/note="Sabrina LTR-2"

LTR
31781..33400
/note="Sabrina LTR-3"

LTR
33897..34585
/note="Sukula LTR-1"

LTR
34586..36407
/note="BARE-1 LTR-3"

LTR
36618..38442
/note="BARE-1 LTR-4"

LTR
38443..40171
/note="Sukula LTR-1"

LTR
40261..41734
/note="similar to tnp2"

LTR
44020..44979
/product="lonoplast intrinsic protein 1"

LTR
44020..44979
/gene="TIP1"

LTR
44020..44979
/product="lonoplast intrinsic protein 1"

LTR
44020..44979
/gene="TIP1"

LTR
44020..44979
/product="lonoplast intrinsic protein 1"

LTR
44020..44979
/gene="TIP1"

LTR
44020..44979
/product="lonoplast intrinsic protein 1"

LTR
44020..44979
/gene="TIP1"

LTR
44020..44979
/product="lonoplast intrinsic protein 1"

LTR
44020..44979
/gene="TIP1"

LTR
44020..44979
/product="lonoplast intrinsic protein 1"

LTR
44020..44979
/gene="TIP1"

LTR
44020..44979
/product="lonoplast intrinsic protein 1"

LTR
44020..44979
/gene="TIP1"

LTR
44020..44979
/product="lonoplast intrinsic protein 1"

LTR
44020..44979
/gene="TIP1"

LTR
44020..44979
/product="lonoplast intrinsic protein 1"

LTR
44020..44979
/gene="TIP1"

LTR
44020..44979
/product="lonoplast intrinsic protein 1"

LTR
44020..44979
/gene="TIP1"

LTR
44020..44979
/product="lonoplast intrinsic protein 1"

LTR
44020..44979
/gene="TIP1"

LTR
44020..44979
/product="lonoplast intrinsic protein 1"

LTR
44020..44979
/gene="TIP1"

LTR
44020..44979
/product="lonoplast intrinsic protein 1"

/codon_start=1
/product="lonoplast intrinsic protein 2"
/protein_id="AAP9012.1"
/db_xref="GI:9623336"

/translation="MASNLRVHLKHCSPSPSLRSFAEFTSTLEFVAVGSAISARM
LTPDVTSMASSLAVTAVAOSRGLFAAFVIAADVSGGVNPAVFPAFAGGIVPTAI
FYWTQGLSGSLACLVHFLSAGQAVPTTRIAVMTGSAISVGVTPFVTVYVHA
GDPGGRKGLATLALGLVGLVGTACVLAAGSLTGSNMPARSFPAVYSGDFKNO
AVYAGPMGAVALVAVHQLVFPSPADEPLPHSRHSGVETVVV"

47279..47443
/rpt_family="stowaway"

/rpt_type="inverted"

/note="similar to Tal1"

join(56694..56822,58480..58482,58563..58645,59572..59738,
59853..60076,60162..60254)

/gene="Rar1"

/product="Rar1"

<56694..>60254

/gene="Rar1"

join(56694..56822,58480..58482,58563..58645,59572..59738,
59853..60076,60162..60254)

/gene="Rar1"

/note="CHORD containing protein"

/codon_start=1

/product="Rar1"

/protein_id="AAP9012.1"

/db_xref="GI:9623337"

/translation="MSAETERSAAAPAPMRCORIGCDAMEFDDNPNCSCHYHRS
PLEHDMKEMSCCKORSHPFLFIAPGATKHTTEKVTYAVLSNKAIPPSAP
VOSKQGVETFACSRROGFSPDSHDSQRAQPAVNTGNTESVOKSVPEPKKVV
DINPEPVCKNCKGCKTYKEDNDHACDHPDPAVHNRGMKCCDVHVEDEME
IPCTKGMHNAADV"

61060..61170

/rpt_family="stowaway"

/rpt_type="inverted"

63185..65008

/note="BARE-1 LTR-5"

BASE COUNT 17560 a 15026 c 15067 g 18326 t

ORIGIN

repeat_region

LTR

Query Match 75.5%; Score 571.4; DB 8; Length 65979;

Best Local Similarity 86.4%; Pred No. 3.4e-138;

Matches 654; Conservative 0; Mismatches 101; Indels 2; Gaps 2;

1 gaaaagaggttgtagaagctcctgtagaagatatactatcatgttgctacagatt 60

3446 GAAAGAAATATTAAAGCTTGTGAAGCAGAAATATTATCTGCTCATACGATT 3387

61 ggttagtcggtgtagatgcgttctctagaagagagcattacggttgcctaatgata 120

3386 GGTAAAGTCACATGATGCGTAAAGAGGATACACCGCTTCCTAATGATA 3327

121 aggaatgaatgaatcccggaagatatactatcatgttgtagatgaagata 180

3326 AGAATGAATGAATCCCAACAGAGATTGTACTGCTATAGATGCTGATCCGAA 3267

181 aactgaataaagccactagaagaatatactatcatgttgtagatgaagata 240

3266 AACTGAATAAAGCCAGGAACCATATCTCTGCGCTTATGCAAAATGCTAG 3207

241 aaagtgtgcaactcaactcacttgccttctagaagagattatctatcccaaa 300

3206 AAAGTGTGCAACTCACTCACTTGTCTTGAAGAGGTTATTCAGGTTCTCGCAA 3148

301 taccagtgtgcaactcactcactagaagaagaacacacttcaactgccttgccttgc 360

3147 TATCTGCTGCACAACTGATCAGAGAAACCATCTTACCTGCCCTTGCCTGCTT 3088

361 cttaagagtagtgccttgccttgccttgccttgccttgccttgccttgccttgc 420

3087 CTTATGAGAGTAGTGCCTTGCCTTGCCTTGCCTTGCCTTGCCTTGCCTTGCCTT 3029

KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
BASE COUNT
ORIGIN

thale cress.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 49981)
Preuss, D., Copenhaver, G. and Kelch, R.
Plant chromosome compositions and methods
Patent: WO 0055325-A 283 21-SEP-2000.
The University of Chicago (US)
Location/Qualifiers
1. 49981
/organism="Arabidopsis thaliana"
/db.xref="taxon:3702"

Query Match 50.4%; Score 381.6; DB 6; Length 49981;
Best Local Similarity 69.4%; Pred. No. 9,3e-89;
Matches 519; Conservative 0; Mismatches 229; Indels 0; Gaps 0;

3 aaagaaggtgtgaagcctcctgatatgaaggtattatcatcagtggtcctatgagatgg 62
1888 AAGGAGATTGTGAACCTGTGAGAGAGGATATTCTTCGATTCAGATTGAGTATG 1947
63 gtagatccggtgtagatagcgttccctaaagaagagcattacggtgctccctaaatgaag 122
1948 GTTAGCTCTAGTCTAGTATGTTCCCAAGAGAGGAGGTTTACAGTAGTCAAGATGAGAA 2007
123 gatgaatgatcccgagagatgatcactggtctatagatggtgatgctatcaggaag 182
2008 GAGGACTGATCTCTCTGCGAGACATCACAGGCCGCCGATGTCATCATTAATAGAAAG 2067
183 ctgaataagccacatggaagaagatcctcccttgccttcttcttgcacatcagatgaag 242
2068 ATAAAGTCTCTACCGAGAAAGACATTTTCCCTTACCATTTATCATCAGATGTTAGAG 2127
243 aggttgcacaaactcacacactctcttcttcttcttcttcttcttcttcttcccaata 302
2128 ATGTGAGCAATCATTAATATCTATTGTTCTTGATGAGATCATCAAGATCTTTCAGATC 2187
303 ccagtgacacatcgtatcagagagaaacacattcaccgcttgcgttacccttgc 362
2188 CCGATTCATCGAGATGACCGAGAGAAACGACTTTCATCTTATGATGATATTTGCT 2247
363 tatagaagatgcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 422
2248 TATCGAGAAATGATCATTCGGCTTTTGTATGCTCTCTCAACATTTTCAGAGATGATG 2307
423 gctatcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttccgt 482
2308 TCTATCTTTCACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2367
483 taaggtcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 542
2368 TATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2427
543 gacacacatcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 602
2428 GAGAAACATCTAGTTTAAATTTGAGAGAAATGATGATGATGATGATGATGATGATG 2487
603 ttaggacataaattcttgaagaagattgaagtcgataagcgaagtcgaagtcgaagtc 662
2488 CTCGAGACAGAGATTTCTGAGCATGATGATGATGATGATGATGATGATGATGATGATG 2547
663 gacaaaagcattccacacagatatacaaggtataagagttccttcttgcacagtggt 722
2548 ACAAGTCTTCAAGCGCTGATTAATGATTAAGAGGATTCCTTGGACATGCTGGT 2607
723 ttctatagaaggttatcaaatcttca 750

Db 2608 TTCTACAGAGATTCATCAAGACTTCA 2635

RESULT 4
AC007209
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS

AC007209 54708 bp DNA PLN 05-APR-2000
Arabidopsis thaliana chromosome II section 74 of 255 of the
complete sequence. Sequence from clones F15011, F1404, T26C18.
AC007209 AE002093
AC007209.5 GI:6598713
HTG.
thale cress.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 54708)
Lin, X., Kaul, S., Rounsley, S.D., Shea, T.P., Benito, M.-I., Town, C.D.,
Feldblyum, T.V., Buell, C.R., Ketchum, K.A., Lee, J.J., Ronning, C.M.,
Koo, H., Moffat, K.S., Cronin, L.A., Shen, M., VanAken, S.E., Umayam, L.,
Tallon, L.J., Gill, J.E., Adams, M.D., Carreira, A.J., Creasy, T.H.,
Goodman, H.M., Somerville, C.R., Copenhaver, G.P., Preuss, D.,
Nierman, W.C., White, O., Eisen, J.A., Salzberg, S.L., Fraser, C.M. and
Venter, J.C.
Sequence and analysis of chromosome 2 of the plant Arabidopsis
thaliana
Nature 402 (6763), 761-768 (1999)
20083487
10617197
2 (bases 1 to 54708)
Lin, X.
Direct Submission
Submitted (09-MAR-2000) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA
On Dec 17, 1999 this sequence version replaced gi:4733964.
The sequence and annotation of chromosome 2 were merged from those
of the individual clones on this chromosome after removing
overlaps. For detailed information, please see the TIGR web site
(http://www.tigr.org/tdb/at/le.html).

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Genes were identified by a combination of three methods: Gene
prediction programs including GRAIL
(ftp://arthur.egr.msu.edu/pub/grail/), GeneFINDER (Phil Green,
http://arthur.egr.msu.edu/pub/grail/), GENSCAN (Chris Burge,
University of Washington), GENSCAN (Chris Burge,
http://www.cds.stu.dk/services/NetGene2/), searches of the
complete sequence against a peptide database and plant EST
databases at TIGR, and manual curation based on those analyses.
Annotated genes are named to indicate the level of evidence for
their annotation. Genes with similarity to other proteins are named
after the database hits. Genes without significant peptide
similarity but with EST similarity are named as 'unknown' proteins.
Genes without protein or EST similarity that are predicted by two
or more gene prediction programs over most of their length are
annotated as 'hypothetical' proteins. Genes encoding tRNAs are
predicted by tRNAscan-SE (Sean Eddy,
http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats were
identified by RepeatMasker (Arian Smit,
http://ftp.genome.washington.edu/RM/RepeatMasker.html). Genes are
numbered from the top to bottom of the chromosome.

We thank the CSHL/Mashu/ABI consortium for sequencing BAC clones
F6P23, F5J6, T17A5, and T13L6, the ESSA group for sequencing clone
F13B4, and Scott Jackson, Jiming Jiang, Klaus Meyer, Eric Richards
and Satoshi Iwata for helpful assistance. In addition, we would
like to thank the TIGR Bioinformatics Department, especially Lixin
Zhou, Hanif Khalek, Michael E. Heaney, Lily Fu, Feng Liang, Jeremy
Peterson, Michael Holmes, and Delwood Richardson for software and
database support.

This work was supported by the National Science Foundation,
Department of Energy and the US Department of Agriculture.

FEATURES		Address all correspondence to: atetigr.org.
source	Location/Qualifiers	
misc_feature	1..54708	/organism="Arabidopsis thaliana"
	/db_xref="taxon:3702"	/chromosome="11"
misc_feature	complement(1..434)	/note="Sequence from clone F15011"
	complement(435..54507)	/note="Sequence from clone F1404"
gene	complement(435..1687)	/gene="F1404.9"
	/note="F1404.9"	/pseudocodon_start=1
mRNA	complement(<4744..>5949)	/gene="At2g13350"
	complement(<4744..>5949)	/gene="At2g13350"
gene	/note="F1404.8; identical to hypothetical protein GB:AD20424"	
	complement(4744..5949)	/gene="At2g13350"
CDS	/note="hypothetical protein"	
	/codon_start=1	/protein_id="AAD28670.1"
mRNA	/db_xref="GI:4733989"	/translation="MSFLAPQQLLEINIIAQLAPVAPRCMTYATAMIDPERLTLRVDNMGTSPTMNDKFFRDLDEALYDSTVIVETIYLHMKDILHVGTVQALISDLVS
	/db_xref="GI:4733989"	PSAMRPTVEVLRLASGRPLGLNIAVGLIDNSQSPLEEDLLEFKHNLISKPG
gene	complement(7452..9489)	LRKSDTSSVDSPPKRVQSVSTNSGDKDESSDSQMVYEPQRTNPMROT
	/note="F1404.7"	KQDLVGTGPMRPRKTAATYPRKNTIEGTGPMRSPRIYITESDLGPSAAVAQIAKAK
CDS	complement(7636..7773,7852..8062,8182..8548,8632..9121)	ALUGKAESTVIVSGERSEVGLRSKLERMOANLPLVDGSSYQPSDSDYTKNSNEN
	/gene="At2g13360"	SSKPNELIVPRNPOMIGAPIQKPSGRNKGSGDGLFCFSCINICIGESIVCGSGSQK
mRNA	complement(7452..9489)	AAKKKKKK
	complement(7452..9489)	8632..9180,9454..9489)
gene	complement(7452..9489)	/gene="At2g13360"
	/note="F1404.7"	complement(7452..9489)
CDS	complement(7636..7773,7852..8062,8182..8548,8632..9121)	complement(7636..7773,7852..8062,8182..8548,8632..9121)
	/gene="At2g13360"	/codon_start=1
repeat_region	/product="alanine-glyoxylate aminotransferase"	
	/protein_id="AAD28669.1"	/db_xref="GI:4733989"
repeat_region	/translation="MDYMGGRPHNLFPVGRVNIPEPVIRAMRNNDYRSPALPALT	
	/db_xref="GI:4733989"	KTLLEPVKKIKFTSTSTPFLPTTGAMESALNTLSPEDRIVSELIQGSFLMWDQ
gene	complement(12400..13193,13358..13451,13543..13896,13995..14176,14264..14441,14540..14695,14939..15184,15294..15374,15643..15822,15902..15988,16358..16468,16558..16800,17381..17587,17747..17810,17888..18012,18110..18203,18561..18628,18716..18874,19158..19319,19451..19538,19652..19793,20166..20229,20559..20610,20779..20882,21011..21070,21364..21447,21558..21782,22156..22575,22795..22862,23018..>23762)	
	/gene="At2g13370"	complement(12400..>23762)
gene	complement(12400..>23762)	/gene="At2g13370"
	/note="F1404.6"	/note="F1404.6"
repeat_region	complement(12500..12464..12500)	
	/rpt_family="TA)n"	

CDS

complement(12820..13193,13358..13451,13543..13896,13995..14176,14264..14441,14540..14695,14939..15184,15294..15374,15643..15822,15902..15988,16358..16468,16558..16800,17381..17587,17747..17810,17888..18012,18110..18203,18561..18628,18716..18874,19158..19319,19451..19538,19652..19793,20166..20229,20559..20610,20779..20882,21011..21070,21364..21447,21558..21782,22156..22575,22795..22862,23018..>23762)

/gene="At2g13370"

/codon_start=1

/product="SNF2 family (CHD1 subfamily) chromodomain protein"

/protein_id="AAD28668.1"

/db_xref="GI:4733988"

/translation="MAEPRNSNDTVSHNVLDENEERQNAATFQSPPLNEDVDGYSE

RFLMNDVVOYOSDPEPGCSIRQNETVADVAVPDVDSYSTRKLTGVMGCTFW

KDQPMQREGSDPAKDSQSGYKEAYHEDHNSKRSKLESDENNDENEDENEM

KOSQADVPADEMILSDYEYQDDESDHVKYKSNPTSRSLPKGSAVHSNRT

SRAIHKNTHSNSNDHGDADMDYEEDEDDPDAPEFYDAADDCGASKHOGM

DVSDDEDEDEIDSDYEDYGTAKPKVROOSGFRKSSAGLEKSFVSSRKRKT

SYODDSEDESDNDEGFRSLARGLTLRQNGSRNTIGQSSERSTSVKVS

VESEDESDIDGKRNKQKDI EEDDAVIERVLMHQLKGEDVQTNKSTVPLVS

QLPDEPDNEMEFLLKMGQSHLCQMTLSDQNLGSGFKVLYNTKVEEIRYT

ALSEELEVDNVSKEMDLITKONSQVRIIADRSKGLGDVPEVTKWOGLSYAE

ATMREVDVIAFAOVAIDYKAREVSIAGCGEQCRKGGKGENSPNAELMLPSVA

SLRKLDEPEMLIGTLRDYDLEGNTLVNSLNDNTVILADENGLGTVQSVMLG

LQNTQQLPGLVPLVPLSTLANMKKEFKMLVGNHIIYVATVSRASRENTNDVHKV

GRPLKFNALITTVYEVVLKDAVLSKIKVYLMVGNHIIYVATVSRASRENTNDVHKV

LLITGTPQNSVEBELMALHFLDGLGKFNKQYEVENKLSFSESLANLHLEPLH

ILRRVITDVEKSLPPIKTERILRVEMSPLOKQYKVIERNHDLNKGVRGNSVLLI

LVLEKCCNRPDLIESADHYGSDIINNSKDKITLSGKVIYDKLILVIREKHV

LIFSOMRMHLLAELVSLRQFORPDRGSKALROAMDHFAAPSDDCFLLSTR

AGGICNLATADTYITDSQNPQNDQAMSKAIRIQOEVENIYRVTSVSEEL

ERRAKRVLDHVLQALNABERLREKTKGNSDKNLSLIFGAELEKEDKNE

ESKRRLVSMIDITELERADEVKTEKTEITELGAKRVNFCNAEDGSGWSMIP

DSVYTAEBALAPRAARNTKSYVDSHPDRISRKKESEPEHTEORRKTEYFV

STPLLESTAQVRMSYGNLPRDAQGFRTYRMFGHNMOCIAEYGVGAAPRE

AOVLEFPLALIDKESEVETGNEPKGVLFEGVPAVELIKVQGLQSLKISIX

NDPISORVLSYLRKPSWMSKGCNNQDIDARLLGLIYHFGNMEKIRLDSGLTAK

IAPEVDHHEFLPRANLKEKATALLMELEAAGKNTAKSRKSKVKDNLINQ

KAPARDKRSQGANVSLSTRDGPRTKAPLEIVREEMSDSDGVDFGFKOKM

EMCEDVLADEIKTLGLRLQTTTADIKPKVLEIRYLEIIGRRIDALVLEHEEDL

YKODRMTRMLNMYVSTPNSGLRDLNIYKLEQKEEGEVGSGNSGRNRQOK

FKTAGNSQSGOQVHKGDITAKFEAKRRRRTENDVQERTTNSNLGLTGCPDLR

repeat_region

repeat_region

gene

misc_feature

gene

Query Match

Best Local Similarity

Matches 522; Conservative 0; Mismatches 234; Indels 0; Gaps 0;

2

62

Db

Qy

Db

Qy

122

Qy


```

repeat_region /note="F14G16"
1528..1684
repeat_region /note="F14G16"
1561..1647
repeat_region /note="F14G16"
1646..1907
repeat_region /note="F14G16"
1672..1814
repeat_region /note="F14G16"
1824..2012
repeat_region /note="F14G16"
1901..1977
repeat_region /note="F14G16"
1975..2479
repeat_region /note="F14G16"
1996..2315
repeat_region /note="F14G16"
2213..2530
repeat_region /note="F14G16"
2465..2565
repeat_region /note="F14G16"
2474..2791
repeat_region /note="F14G16"
2793..2826
repeat_region /note="F14G16"
2826..3601
repeat_region /note="F14G16"
3602..4665
repeat_region /note="F14G16"
5136..5243
repeat_region /note="F14G16"
5241..5443
repeat_region /note="F14G16"
5887..6013
repeat_region /note="F14G16"
6002..6401
repeat_region /note="F14G16"
7139..7282
repeat_region /note="F14G16"
7300..7710
repeat_region /note="F14G16"
7701..7747
repeat_region /note="F14G16"
7738..7837
repeat_region /note="F14G16"
7835..8136
repeat_region /note="F14G16"
8114..8270
repeat_region /note="F14G16"
10255..10398
repeat_region /note="T32N15_del_retroTn"
10261..10435
repeat_region /note="T32N15_del_retroTn"
10424..10834
repeat_region /note="T32N15_del_retroTn"
10824..10865
repeat_region /note="T32N15_del_retroTn"
10824..10870
repeat_region /note="T32N15_del_retroTn"
10860..11259
repeat_region /note="T32N15_del_retroTn"
10860..10959
repeat_region /note="T32N15_del_retroTn"
10958..11259
repeat_region /note="T32N15_del_retroTn"
11231..11812
repeat_region /note="T32N15_del_retroTn"
11238..11394
repeat_region /note="T32N15_del_retroTn"
11820..11906
repeat_region /note="T32N15_del_retroTn"
11906..12167
repeat_region /note="T32N15_del_retroTn"

```

```

repeat_region 12162..12238
repeat_region /note="T32N15_del_retroTn"
12237..12741
repeat_region /note="T32N15_del_retroTn"
12237..12741
misc_feature complement(12310..15735)
polypeptides /note="pseudogene; similar to retrovirus-related
12735..13052
repeat_region /note="T32N15_del_retroTn"
13053..13086
repeat_region /note="T32N15_del_retroTn"
13087..13862
repeat_region /note="T32N15_del_retroTn"
13864..14927
repeat_region /note="T32N15_del_retroTn"
14920..15027
repeat_region /note="T32N15_del_retroTn"
15024..15226
repeat_region /note="T32N15_del_retroTn"
15224..15350
repeat_region /note="T32N15_del_retroTn"
15340..15739
repeat_region /note="T32N15_del_retroTn"
21183..22195
gene /gene="F14G16.2"
join(21183..21320,21398..21662,21795..21897,21973..22195)
CDS /gene="F14G16.2"

```

Query Match 50.4%; Score 381.6; DB 8; Length 63292;
 Best Local Similarity 69.4%; Pred. No 9.6e-89;
 Matches 519; Conservative 0; Mismatches 229; Indels 0; Gaps 0;

```

QY 3 aaagaagttggaagctccctgatagaaggtatatactatcatgtgtgcatagcgtatg 62
Db 49198 AAGGAGATTGGAAGCTGTTGGAAGCAGGATATATCTTCGATTGATGATGATGAT 49257
QY 63 gtgagtcggtgcatagcgttccctgaagaaggagagcgttgcgttgcctaatgaa 122
Db 49258 GTTAGTCAGTTCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 49317
QY 123 gatgaattgataccgcgagagattatcatcactgataagagtggtgcatgaa 182
Db 49318 GACGAGCTGATTCCTACTCGACATCACAGCCCGGATGATGATGATGATGATGAT 49377
QY 183 ctgaataagcccaactgaagaagatcatcacccttgccttgccttgccttgccttgc 242
Db 49378 ATGATGCTGCTACACGAGAAACGACATTTCCCTTACCATTTATGATGATGATGAT 49437
QY 243 aggttgcacactcaacacttgccttgccttgccttgccttgccttgccttgccttgc 302
Db 49438 ATGTTAGCAATCATTAATATGTTGTTCTTCTTGTATGATGATGATGATGATGAT 49497
QY 303 ccagttgcacactgataagagagaaacacacttgccttgccttgccttgccttgccttgc 362
Db 49498 CCGATTTCATCGAGATACCAAGGAGAAACGACCTTCACTTCACTTCACTTCACTTCACT 49557
QY 363 tatagaagatgacttgccttgccttgccttgccttgccttgccttgccttgccttgccttgc 422
Db 49558 TATCGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 49617
QY 423 gctatattctcaacttgccttgccttgccttgccttgccttgccttgccttgccttgccttgc 482
Db 49618 TCTATCTTACACATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 49677
QY 483 tacggctcttcttgccttgccttgccttgccttgccttgccttgccttgccttgccttgccttgc 542
Db 49678 TATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 49737
QY 543 gacacaaactctgtccttgccttgccttgccttgccttgccttgccttgccttgccttgccttgc 602
Db 49738 GAGAAACATCTAGTTTAAATTGGAGCAATGTCATCTTATGATGATGATGATGATGAT 49797
QY 603 ttagacataaattctcgaagaaggtattgaagtcgataagcgaagttgatac 662

```

[illegible]

	repeat_unit	12162..12238 /note="T32N15_del_retroTrn"
	repeat_unit	12237..12741 /note="T32N15_del_retroTrn"
	misc_feature	12310..15735 /note="pseudogene, similar to retrovirus-related polypoteins"
	repeat_unit	12735..13052 /note="T32N15_del_retroTrn"
	repeat_unit	13053..13086 /note="T32N15_del_retroTrn"
	repeat_unit	13087..13862 /note="T32N15_del_retroTrn"
	repeat_unit	13864..14927 /note="T32N15_del_retroTrn"
	repeat_unit	14920..15027 /note="T32N15_del_retroTrn"
	repeat_unit	15024..15226 /note="T32N15_del_retroTrn"
	repeat_unit	15224..15350 /note="T32N15_del_retroTrn"
	repeat_unit	15340..15739 /note="T32N15_del_retroTrn"
	repeat_unit	21183..22195 /gene="AT4g07310"
CDS	join(21183..21320,21396..21662,21795..21897,21973..22195)	/note="contains similarity to Arabidopsis thaliana hypothetical proteins, see GB:AF058826 similarity to"
gene	/protein_id="CA881101.1" /db_xref="GI:7267327" /translation="MESFPLSLIKPSIRGWCIRGRVRFELNSLVSSKMGILLAE EHGMTEATWGYKMSDHYKDFINEGEVITTFNGVENSQSVRATHTSPFIGEVDIV LFTKISDENREVRLECLATYKKEALDPDHNRRTYGGVIAVLGWKKIDRYEDGKNRVY CTAGISTVPEPDPIPESNEIHEI"	
exon	21183..21320 /gene="AT4g07310" /number=1	
Intron	21321..21397 /gene="AT4g07310"	
exon	21398..21662 /gene="AT4g07310" /number=2	
Intron	21663..21794 /gene="AT4g07310"	
exon	21795..21897 /gene="AT4g07310"	
Intron	21898..21972 /gene="AT4g07310"	
exon	21973..22195 /gene="AT4g07310"	
misc_feature	23319..28183 /note="pseudogene, similar to retrovirus-related polypoteins"	
exon	complement(29065..29454) /gene="AT4g07320"	
gene	29065..31484 /gene="AT4g07320"	
gene	complement(join(29065..29454,29519..29703,30603..30969,	

OY	3	aaagaggttggaagctctcttgagagaagtatactatcatgatgtgtctcaagcattg	62
Db	49198	ANGGAGATTGTGAAACCTTTGTGAAGCAGAGATATATCTTCGCAATTCAGATAGTACTTG	4925
OY	63	gttagtcggttgatagtcgtttactaagaagaagacatacgcattgacctaatgtatg	122
Db	49258	GTTACTCTAGTTCATGTAGTTCGCCAAGAAAGAGAGTTCACATAGTCAAGATGTGAAA	49317
OY	123	gatgaattgatcccgcsagagattatctactcgtctatagaatgtgtatcgattccagaa	182
Db	49318	GACGAGCTGATCTCTACTCGGACAAATCACAGGCCGCCGATGTGCATCATATATGAAG	49377
OY	183	ctgataaagccactaagaaagacatcatcactcttgctttatcgaccatctgtcagaa	242
Db	49378	ATGAATCTCGTCACAGGAAGACCATTTTCCTTACATATATGATAGATGTGTAGG	49437
OY	243	aggctgtccaacacacacactcttcctttctagaagcgttatctagtctcccaata	302
Db	49438	ATGTTTACCAATATATATATATCTATTTCTTCTGTGATGGATCTCAAGATTTCTTCAGATC	49497
OY	303	ccagcttgacaatctgatataagaagaacaacttcaactgcctcttgtaaccttgc	362
Db	49498	CCGATTTCATGAGATGATCACCAGGAAACGACTTTTCACCTGCTTTATGATACATTGCT	49557
OY	363	tatagaagtatgccttttgcttatgtaatgtacactgcaccacttcaaaagatgtatg	422
Db	49558	TATCGAGAAATGTCAATTCGGCTCTTTGATATGCTCTCGACAACTTTCACAGAGTATGATG	49617
OY	423	gctatatctcttaacttttgtaaaaatattgtcgaagcttccatgaagatcttcgct	482
Db	49618	TCTATCTTCACAGACATGATGATGAGATATCATGTGAAGTTTCAATGATGATTTTCAGTT	49677
OY	483	tacgggtctctcttltgatgatgtgcctcgaaccttgcacacttgcagatctcaagagatgaa	542
Db	49678	TATGATGCTCATTTGTAGAGATGCTTTAGCGAATCTTACACAGAGTGTAGCAAGATGTGAG	49737
OY	543	gacaccaatctgtcttgaatgagggagagagccacttaagttatggttaagaagcatcgtc	602
Db	49738	GAGAAACATCTTAGTTTTAAATTTGGAGAAATGTCACTTTATGTGCGACAGATGGATATGTT	49797
OY	603	ttagagacataaaattcttgaagaaggtatltgaagtcatagaagctaaagttatgtcaatc	662
Db	49798	CTCGACACACAGATTTCTTGACACATGGTATGTGAATGTGATGAGCTTAAGATTGAGGTATAG	49857
OY	663	gacaaaatgtcatcccccacacagatatalaaaggtataaagaagcttctctgtcatgtgt	722
Db	49858	ACAAGTCTTCAAGGCCCTGATATATGTTAAGTATATGAAGGTTTCTTGGACATGCTGCT	49917
OY	723	tctcatagaagttatcaagaatttca 750	
Db	49918	TTCTACAGGATTCATCAAGACTTCA 49945	

RESULT	7				
LOCUS					
DEFINITION					
ACCESSION					
VERSION					
KEYWORDS					
SOURCE					
ORGANISM					
REFERENCE					
AUTHORS					
AC007187	53685 bp	DNA	PLN	05-APR-2000	
Arabidopsis thaliana chromosome II section 63 of 255 of the complete sequence. Sequence from clones T1806, F7E22, F23M2.					
AC007187	AC002093				
AC007187.4	GT:6598701				
HTG.					
thale cress.					
Arabidopsis thaliana					
Eumetazoa; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.					
Lin, X., Kaul, S., Rounsley, S.D., Shea, T.P., Benito, M.-I., Town, C.D., Fujii, C.Y., Mason, T.M., Bowman, C.L., Barnstead, M.E., Feldblyum, T.V., Buell, C.R., Ketchum, K.A., Lee, J.J., Ronning, C.M., Koo, H., Moffat, K.S., Cronin, L.A., Shen, M., VanAken, S.E., Umayam, L., Tallon, L.J., Gill, J.E., Adams, M.D., Carrera, A.J., Creasy, T.H., Goodman, H.M., Somerville, C.R., Copenhaver, G.P., Preuss, D.,					

Nierman, W.C., White, O., Eisen, J.A., Salzberg, S.L., Fraser, C.M. and Venter, J.C.
 Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*
 Nature 402 (6763), 761-768 (1999)
 JOURNAL MEDLINE 20083487
 PUBMED 10617197
 REFERENCE 2 (bases 1 to 95685)
 LIn.X.
 Direct Submission
 Submitted (09-MAR-2000) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA
 On Dec 17, 1999 this sequence version replaced gi:4733998.
 The sequence and annotation of chromosome 2 were merged from those of the individual clones on this chromosome after removing overlaps. For detailed information, please see the TIGR web site (<http://www.tigr.org/tcdb/at/at.html>).

Genes were identified by a combination of three methods: Gene prediction programs including GRAL (<http://artur.epm.ornl.gov/pud/xgral/>), Genefinder (Phil Green, University of Washington), Genscan (Chris Burge, <http://genomic.stanford.edu/GENSCAN.html>), and NetPlantGene (<http://www.cbs.dtu.dk/services/NetGene2/>), searches of the complete sequence against a peptide database and plant EST databases at TIGR, and manual curations based on those analyses. Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as 'unknown' proteins. Genes without protein or EST similarity, that are predicted by two or more gene prediction programs over most of their length are annotated as 'hypothetical' proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats were identified by RepeatMasker (Arian Smit, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>). Genes are numbered from the top to bottom of the chromosome.

We thank the CSHL/Mash/ABI consortium for sequencing BAC clones F6P23, F5J6, T17A5, and T13L16, the ESSA group for sequencing clone F13D4, and Scott Jackson, Jiming Jiang, Klaus Meyer, Eric Richards and Satoshi Tabata for helpful assistance. In addition, we would like to thank the TIGR Bioinformatics Department, especially Lixin Zhou, Hanif Khailak, Michael E. Heaney, Lily Fu, Feng Liang, Jeremy Peterson, Michael Holmes, and Delwood Richardson for software and database support.

This work was supported by the National Science Foundation, Department of Energy and the US Department of Agriculture.

Address all correspondence to: atetigr.org.

FEATURES
 source Location/Qualifiers

1..95685
 /organism="Arabidopsis thaliana"
 /cultivar="Columbia"
 /db_xref="taxon:3702"
 /chromosome="II"
 <1..301
 /note="Sequence from clone T1806"
 complement(302..95484)
 /note="Sequence from clone F7E22"
 344..2412
 /gene="At2g11670"
 /note="F7E22.16; TNP2 protein of the Tam1 family of transposons [Antirrhinum majus]"
 /pseudo
 373..434
 /rpt_family="POLY_A"
 3213..3703
 /note="DNA sequence related to Athila retrotransposon"
 <3898..>4803
 /gene="At2g11680"

gene
 <3898..>4803
 /gene="At2g11680"
 /note="F7E22.15"
 3898..4803
 /gene="At2g11680"
 /codon_start=1
 /product="putative retroelement pol polyprotein"
 /protein_id="A028680.1"
 /db_xref="GI:4734002"
 /translation="MYREGIVGHTSKKGVYAKVAVMMLOLPKTKDIRSFLG HAGRFRFIDEESKLARPLRLRCKRAEPEEECTAFKSTKALVASPIYOAPRD YPEELMCDSDVAVGAVLGOIKDRKLHVIYASGTDPTVDRATTEKELLAVAFEF KFSYLVASVYVYTDHAAIRHLYAKDKRPRLRLNIILOEDMEIVKKGLENVA DLSRMRIDEVEPIDDSMEPEQLMAIQOLNESAOINKSLDQVCAYEDKLMYADHVKY LFGSEPNLSYEKKFEDINHF"
 complement(join<7068..7358,7445..7738,7824..8150,8234..>8380))
 /gene="At2g11690"
 complement(<7068..>8380)
 /gene="At2g11690"
 /note="F7E22.14; related to En/Spm transposon family of maize"
 complement(join(7068..7358,7445..7738,7824..8150,8234..8380))
 /gene="At2g11690"
 /codon_start=1
 /product="En/Spm-like transposon protein"
 /protein_id="A028679.1"
 /db_xref="GI:4734001"
 /translation="WGRGVATKRAFLQARDAYOKLEAOQALLINRINDONEVGL AVGNKSDVSISETSKVATRCQLDWSNEDIIVGBSEFCSABRTYKIGTIPKNAAY IYVFEKEDAPVWRPVATFISLGNMGTIIPWSPNKITVDNANVSPTEHNTAGSSAP PDRLLIDMNVVEDIIVAGHLISDPKELANNIPGLPNAALVYKVKIFPKDAYLRP TSEMILMGDAHELIIWPIQRIQISGNSPTSPESPASCSMAKTTKSNDRKRL LDGSGSEIYAEGRICSTNPBDKHFPLPDPASKVWVYLKIGSAKVMRPNSEVHFV GDALGTVAMPNDRLRPM"
 complement(810..14100)
 /note="putative retroelement. Structure: LTRa, retroelement, LTRB"
 9615..11655
 /gene="At2g11700"
 /note="F7E22.13"
 /pseudo
 complement(11014..11053)
 /rpt_family="(CAAAA)n"
 12552..12623
 /rpt_family="POLY_A"
 14100..14907
 /note="DNA sequence similar to Athila retrotransposon"
 complement(15034..16032)
 /gene="At2g11710"
 /note="F7E22.18"
 /pseudo
 complement(16701..19304)
 /gene="At2g11720"
 /note="F7E22.17; TNP2 protein of the Tam1 family of transposons [Antirrhinum majus]"
 /pseudo
 19615..21837
 /gene="At2g11730"
 /note="F7E22.12"
 /pseudo
 complement(22231..23920)
 /gene="At2g11740"
 /note="F7E22.11; TNP2 protein of the Tam1 family of transposons [Antirrhinum majus]"
 /pseudo
 complement(25972..31340)
 /note="putative retroelement. Structure: LTRa, retroelement, LTRB"
 26459..29209
 /gene="At2g11750"
 /note="F7E22.10"

Query Match	50.3%;	Score 380.8;	DB 8;	Length 95685;
Best Local Similarity	69.1%;	Pred. NO. 1.6e-88;		
Matches 520;	Conservative 0;	Mismatches 232;	Indels 0;	Gaps 0

QY	3	aaagaagttgtgaaagctcctctgtagaaggatattatctacatggttgcataagcgaattg	62
Db	33841	AACGGAATTATATGAAACCTGTTGGAGACGAGGATTTATCTATCCATTTACAGATGACAGTTGG	33900
QY	63	gtgaagccggtgcatagagcttccctaagaaggagagcatcactggtccctaatgataag	122
Db	33901	ATTAGGCCAGTTTCATGTACTGTACTTAAGAAAGGAGGGGTTCACGTGATCAAGATATAGAAA	33960
QY	123	gatgaattgatcccgacagaagattlccacttggtcgtatagatgtgtatgcattacgaaga	182
Db	33961	GACGAGCTGATTCCTACTTCGACAMTACACAGGCATCGGATGTGCATGTATTCAGCAAAAG	34020
QY	183	ctggaataaagccactcagaagaagatatattacccttggccttattatgcacatatgctagaa	242
Db	34021	CTGAATTCCTGCTACACAGGAAGAACCATTTTCCCTTACCATTTATGTGATTCAGATGTTTGGAG	34080
QY	243	aggtgtgccaaactacacacactccgcttcttcagaaggtatctcagtttctccccaata	302
Db	34081	AGGTTAGCAAAATCATACAGATATTATTGTTCCCTTGATGATACATCATTTCTTTAAATC	34140
QY	303	ccagttgcacaactcgtatccagagaagaacaacttccacttcccttgcgttacccttgc	362
Db	34141	CCATATTCATCCAGATGACACAGAGAAACCACTTTCACCTGCCCTATGATGATCAATTTACT	34200
QY	363	tatagaagtatgccttttggcttgatgtaatgcacttgcacacttccaaagtgtatgata	422
Db	34201	TATTGAGAAATGCCCTTTGGCTTTGTATATCTCTTTCGACAATTTTAAAGAGGATGTGATG	34260
QY	423	gctatatctcctaactcttcttggaataatgtgcgaagtttcatatgatgactttccgtt	482
Db	34261	TCTATCTTCATAGACATGATTGATGAGACATCATGTGAAGTTTTCATGATGATATTTTCAGGT	34320

RESULT	8				17-JAN-2002
AX059456/c					
LOCUS	AX059456	40480 bp	DNA	PAT	
DEFINITION	Sequence	189	from Patent	W00055325.	
ACCESSION	AX059456				
VERSION	AX059456.1	GI:12311561			

KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source

thale cress.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta
Spermatophyta; Magnoliophyta; eudicotyledons: core eudicots,
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
1 (phases 1 to 40480)
Preuss,D., Copenhaver,G. and Keith,K.
Plant chromosome compositions and methods
Patent: WO 0055325-A 189 21-SEP-2000;
The University of Chicago (US)
Location/Qualifiers
1. 40480

BASE COUNT	ORIGIN				
12306	a	7912	c	8765	g
				11297	t
				200	others

Query Match	50.2%;	Score 380;	DB 6;	Length 40480;
Best Local Similarity	68.9%;	Pred. No. 2.4e-88;		
Matches 521;	Conservative	0;	Mismatches 235;	Indels 0;
				Gaps 0

QY	2	aaagaggtgtgtgaagctccctcgatbaaggtattactaatagttgcacatagsgatg	61
Db	32944	AAAAGAAATTTTGAACCTCCTTGAGGCTGGTGTCATCTTCTCATCTTGATGATGTCCTTG	32885
QY	62	ggtgagtcgcgtgtgacaagcgtttccctaagaaggagagcattacggtgtgtcccatatga	121
Db	32884	GGTTTCTCCAGTGCATTTGGCTCCCTAAAAAGGGTGGAATGACGTGTGTCAAAAATGAAA	32822
QY	122	gataaatgtatcccccagagagatatactacgtgcataagatgtagtgcattcagaagaa	181
Db	32824	ACATGAACTAAACCTCCTACTAGACATATACTGGTCATAGAAAAGTGCAATTGATTATAGGAA	32765
QY	182	actgataaagcgcattagaagaatcatctacccttgcctttatagccatatactaga	241
Db	32764	GTTTAATGCTGCATCTAGGAAAGATCATTTTCTTTACCATTCAATTGACCAAAATGCTTGA	32705
QY	242	aaggtttgccaaactcacacaacattctgtctttctagaagttatctctagtcttcccaaat	301
Db	32704	ACGTTTACTATATCAVCCATACTATTGCTTCTTGATGATATATGTTGTTCTTCCAAT	32645
QY	302	accggttgcacatctgtatcagagagaaacaacatttcacctgccttctcgtttacattgc	361
Db	32644	ACCATTTCACCTTAATGATTCAGAAAAAACACCTTTCACATGTCCCTACGGAACTTTTGC	32585

QY 362 ttatagacgtatgctcttggctatgtatgacacgtcccttcaagaatgatgat 421
 Db 32584 TTATAGAGAAATGTCATTTGGTTTATGTAATGCTCCCAACATTTGAGAGTGATAC 32525
 QY 422 ggcctatctccttaactcttggaataatgctgaggttctcaagtgacttccgt 481
 Db 32524 CTCTATATTTTTCAGATTTTAATGAGAGAGATGCTGAGGTTTTCATGACGATTTTCCGT 32465
 QY 482 ttaaggctctcttcttgatgtatgctcgaacacgttccagcttcttcaagatgttaa 541
 Db 32464 CTATGGCCCCCTCTTCTCTCTCATGTTTGTGTAATCTTGACAGGGTACTGACAGGTGCGA 32405
 QY 542 agacacccaactctctctggaatgagggaggaagtgccacttattgttataagagcatcgt 601
 Db 32404 ACAGACGAAATCTTTGCTTCATATGGGAAAGGTCTCATTTTCATGAGGAGCAAGCATAGT 32345
 QY 602 cttaaggacataaattcttgaagaaggtatgtaagtcgataaggttaaggttgcacat 661
 Db 32344 GTTGGGTCCAGATATATCAGAGAGGATATAGAGTTGACAAAGGAAATTTGAAGTAT 32285
 QY 662 cgacaaatgcatcccccacagatatacaaggtataaagttctcttgcgtatgctggt 721
 Db 32284 GATGCGATTGCGACGCCCAAAACGCTGAGAGCATACAGAGTTCTGTCATGCTGG 32225
 QY 722 ttctatagaaggttatacaagaattctcaagaagt 757
 Db 32224 GTTCTACAGAGATTTATTAAGACTTCTCAAGAT 32189

RESULT 9

AC016828 76911 bp DNA HTG 29-SEP-2000
 LOCUS Arabidopsis thaliana chromosome 3 clone IGF-F21A14, *** SEQUENCING
 DEFINITION IN PROGRESS ***, 1 ordered pieces.
 AC016828
 AC016828.5 GI:10337645
 VERSION HTG: HTGS.PHASE2.
 KEYWORDS thale cress.
 SOURCE Arabidopsis thaliana
 ORGANISM Arabidopsis thaliana; Streptophyta; Embryophyta; Tracheophyta;
 Eukaryota; Viridiplantae; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE

AUTHORS

Lin, X., Kaul, S., Town, C. D., Beito, M. I., Creasy, T. H.,
 Bowman, C. L., White, O., Fujii, C. Y., Utterback, L. J., Barnstead, M. E.,
 Arabidopsis thaliana 'IGF' BAC 'F21A14' genomic sequence near
 marker 'unpublished'

JOURNAL 2 (bases 1 to 76911)
 REFERENCE Lin, X. and Kaul, S.
 AUTHORS Direct Submission
 TITLE Submitted (08-DEC-1999) The Institute for Genomic Research, 9712
 JOURNAL Medical Center Dr., Rockville, MD 20850, USA, xline@igf.org
 COMMENT On Sep 29, 2000 this sequence version replaced gi:6692721.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 1 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * provided by the submitter.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 * 1 76911: contig of 76911 bp in length.
 * Location/Qualifiers
 1. 76911
 /organism="Arabidopsis thaliana"
 /db_xref="taxon:3702"
 /chromosome="3"
 /clone="IGF-F21A14"
 BASE COUNT 21697 a 16043 c 16139 g 22185 t 847 others

ORIGIN

Query Match 50.2%; Score 380; DB 2; Length 76911;
 Best Local Similarity 68.9%; Pred. No. 2, 6e-88;
 Matches 521; Conservative 0; Mismatches 235; Indels 0; Gaps 0;

QY 2 aaagaaggtgtagagcttcctgagtgaggtatatactatcaatgctgcatagcatg 61
 Db 16644 AAAAGAAATTTTGAAGCTCTTGAGGCTGCTGTCATCTATCTATCTCTATAGTCTTG 16703
 QY 62 ggtagtcggtgcatagcgttccctaaagaagggagcattaccggttgccttaagtaa 121
 Db 16704 GGTTCGCCAATGTCATTCGCGCCCTAAAGGGGGAATGACGTGTGTCAAAATGAGAA 16763
 QY 122 ggtatgaatgtagcccgagaagattatcaactgctatagatgtagtgcattccgaa 181
 Db 16764 AGATGAAGTAACTTCCCTCTACAGACTTAACTGGTCATGAGATGTGCATTAATATAGGA 16823
 QY 182 actgaataaagccactaggaagaatcatctacccttgccttatacgaccatgctaga 241
 Db 16824 GTTAATCTCTCATCTAGAGAAAGATCATTTTCCCTTAACATTCATTAACCAATGCTTGA 16883
 QY 242 aaggttgcacaactcaacacactctgcttctagacggtatctagtcttcccaat 301
 Db 16884 ACGTTTACCTAATCATCATCACTAATGCTTCTTGATGAGATATAGTGCTTCTTCAAT 16943
 QY 302 accagttgcaacaactgtagcaggaagaacacacttcaacccttgcgttactctg 361
 Db 16944 ACCAATTCACCCCTAATGATCAAGAAACCACTTTTACATGATGCTTCAACGAACTTTTGC 17003
 QY 362 ttatagacgtatgctcttggctatgtatgacacgtccacacttcaagaatgatgat 421
 Db 17004 TTATAGAGAAATGTCATTTGGTTTATGTAATGCTCCCAACATTTGAGAGTGATAC 17063
 QY 422 ggcctatctccttaactcttggaataatgctgaggttctcaagtgacttccgt 481
 Db 17064 CTCTATATTTTTCAGATTTTAATGAGAGATGCTGAGGTTTTCATGACGATTTTCCGT 17123
 QY 482 ttaaggctctcttcttgatgtatgctcgaacacgttccagcttcttcaagatgttaa 541
 Db 17124 CTATGGCCCCCTCTTCTCTCATGTTTGTGTAATCTTGACAGGGTACTGACACAGTGGGA 17183
 QY 542 agacacccaactctctctggaatgagggaggaagtgccacttattgttataagagcatcgt 601
 Db 17184 ACAGACGAAATCTTTGCTTCATATGGGAAAGGTCTCATTTTCATGAGGAGCAAGCATAGT 17243
 QY 602 cttaaggacataaattcttgaagaaggtatgtaagtcgataaggttaaggttgcacat 661
 Db 17244 GTTGGGTCCAGATATATCAGAGAGGATATAGAGTTGACAAAGGAAATTTGAAGTAT 17303
 QY 662 cgacaaatgcatcccccacagatatacaaggtataaagttctcttgcgtatgctggt 721
 Db 17304 GATGCGATTGCGACGCCCAAAACGCTGAGAGCATACAGAGTTTCTTGGTCAAGCTGG 17363
 QY 722 ttctatagaaggttatacaagaattctcaagaagt 757
 Db 17364 GTTCTACAGAGATTTATTAAGACTTCTCAAGAT 17399

RESULT 10

AB046431/C 85962 bp DNA PLN 18-APR-2001
 LOCUS Arabidopsis thaliana DNA, chromosome 5 centromere region,
 DEFINITION clone: T3B16.

ACCESSION AB046431.1 - GI:9971602
 VERSION AB046431.1 - GI:9971602

KEYWORDS

SOURCE

ORGANISM

Arabidopsis thaliana (cultivar: Columbia) DNA, clone: T3B16.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE 1 (sites)
 AUTHORS KumeKawa,N., Hosouchi,T., Tsuruoka,H. and Kotani,H.
 TITLE The size and sequence organization of the centromeric region of Arabidopsis thaliana chromosome 5
 JOURNAL DNA Res 7 (6), 315-321 (2000)
 MEDLINE 21082928
 REFERENCE 2 (bases 1 to 85962)
 AUTHORS Kotani,H. and KumeKawa,N.
 TITLE Direct Submission
 JOURNAL Submitted (21-JUL-2000) Hirokazu Kotani, Kazusa DNA Research Institute, Lab. Chromosome Research II, 1532-3 Yana, Kisarazu, Chiba 252-0812, Japan (E-mail:kotani@kazusa.or.jp, Tel:81-438-52-3920, Fax:81-438-52-3921)
 FEATURES
 source
 1. 85962
 /organism="Arabidopsis thaliana"
 /cultivar="Columbia"
 /db_xref="taxon:3702"
 /chromosome="5"
 /map="centromere"
 /clone="T3B16"
 1. 2000
 /note="overlap to BAC F06024, please refer for analysis and annotation"
 2000. 11119
 /note="similarity to 180 bp centromeric repeat"
 11120. 13109
 13110. 14148
 /note="similarity to 180 bp centromeric repeat"
 14149. 14678
 /note="similarity to 5S rDNA (1 copy) and spacer (1 copy)"
 14679. 16228
 16229. 17527
 /note="similarity to fragments of Athila retrotransposons"
 17528. 18305
 /note="uncharacterized tandem repeat"
 18306. 18360
 18361. 19273
 /note="similarity to fragments of Athila retrotransposons"
 19274. 20452
 20453. 28235
 /note="similarity to fragments of Athila retrotransposons"
 28236. 30040
 30041. 31528
 /note="similarity to fragments of Athila retrotransposons"
 31529. 33152
 33153. 44423
 /note="similarity to fragments of Athila retrotransposons"
 44424. 45933
 /note="uncharacterized tandem repeat (2 copies)"
 45934. 47691
 47692. 55569
 /note="similarity to fragments of Athila retrotransposons"
 55570. 57327
 57328. 61906
 /note="similarity to fragments of Athila retrotransposons"
 61907. 63599
 63600. 64534
 /note="similar to 5S rDNA (2 copies) and spacer (2 copies)"
 64712. 68805
 /note="similarity to 180 bp centromeric repeat"
 68806. 70552
 70553. 81073
 /note="similarity to fragments of Athila retrotransposons"
 81074. 82821
 82821. 85962
 /note="similarity to 180 bp centromeric repeat"
 BASE COUNT 25983 a 16785 c 17774 g 25420 t
 ORIGIN
 Query Match 50.2%; Score 380; DB 8; Length 85962;

Best Local Similarity 68.9%; Pred. No. 2,66-88;
 Matches 521; Conservative 0; Mismatches 235; Indels 0; Gaps 0;
 QY 2 aaaaaggttctgaagctccctgataagagatattatcatatgctgataagagatg 61
 Db 77876 AAAAGAAATTTTGAACAGCTGTGATGCTGTGATCATACCTATACCTGATGATGAT 77817
 QY 62 ggtgagtcggtgataagcgtctcttaagaagagagagatccggttccctaatgata 121
 Db 77816 GGTTCCTCAAGTGTGATGCTGTGCTTAAAGAGGAGATGATGCTGTGCTTAAAGAA 77757
 QY 122 ggaatgattgaccccgagagatattatcagctgctgataagatggtgataagagaa 181
 Db 77756 AGATGAACGTGATACCTACTACACTATACCTGCTGATGATGATGATGATGATGAA 77697
 QY 182 actgaataaagccatgagaaagatcattacccttgcttcttatacgacataagc 241
 Db 77696 GTTGAATGCTGATCATCATATATATGCTTCTTGAATGATGATGATGATGATGATGAA 77637
 QY 242 aaggtgtccaaactcagacacttcgtcttctgagaggtatctagttctcccaat 301
 Db 77636 ACGATTAGCTATCATCATATATATGCTTCTTGAATGATGATGATGATGATGATGAA 77577
 QY 302 accagttgcaaatctgataagagaaacacacttcacccctgctgataagcttgc 361
 Db 77576 ACCAATTCACCCATCATCATATATATGCTTCTTGAATGATGATGATGATGATGATGAA 77517
 QY 362 ttataagcagatgcttctgctgataagatgataagcctgacacttcaagaatgata 421
 Db 77516 CTATAGAGAAATGCTATGCTTCTTGAATGATGATGATGATGATGATGATGATGAA 77457
 QY 422 ggcatactctcaactcttgataaataatgctgaggtttatgataagacttccgt 481
 Db 77456 CTCTATATTTTTCAGACTTAATGAGAGATGAGATGAGATGAGATGAGATGAGATGAG 77397
 QY 482 ttacggctctcttgatgataagctcagacactgacagcttcaagaatgata 541
 Db 77396 CTATGCGCCCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 77337
 QY 542 agacacccaatctgcttctgataagagagagagagagagagagagagagagagag 601
 Db 77336 GGAAGAGAAATGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 77277
 QY 602 ctaagacataaataatctgataagagagagagagagagagagagagagagagag 661
 Db 77276 ATTGGGTCACAAATATCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 77217
 QY 662 cgacaaaatgcataccacagagatatacaaggtataaagagagagagagagagag 721
 Db 77216 GATGACGCTGACGACCAAAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 77157
 QY 722 ttctatagaaggtattatcaaatcttcaagaagt 757
 Db 77156 ATTCTATAGAGATCATTAAGACTTCTTCAAGAT 77121
 RESULT 11
 AB046426/c
 LOCUS Arabidopsis thaliana DNA, chromosome 5 centromere region,
 DEFINITION
 AB046426
 VERSION AB046426.1 GI:9971597
 KEYWORDS centromere,
 SOURCE Arabidopsis thaliana (cultivar:Columbia) DNA, clone:F28N5.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 REFERENCE 1 (sites)
 AUTHORS KumeKawa,N., Hosouchi,T., Tsuruoka,H. and Kotani,H.
 TITLE The size and sequence organization of the centromeric region of Arabidopsis thaliana chromosome 5

4748 CTAAGAAGATGCCATTGGTTATGCATGCTCCGACACATTTCAGAGGTGTATGAC 47487

REFERENCE

JOURNAL
 SUBMITTED (12-JAN-2000) Arabidopsis thaliana Genome Center,
 Department of Biology, University of Pennsylvania, 38th and
 Hamilton Walk, Philadelphia, PA 19104-6018, USA
 5 (bases 1 to 118986)


```

/codon_start=1
/evidence-not_experimental
/product="F7P2.6"
/protein_id="AF24522.1"
/db_xref="GI:6691184"
/translation="MTVALFFPYLGVQUGLPRKVENATSSVQGIYANDLKISCV
HDVLSRPAKRFAPKLELPLEPDAVPSHEMKRFLVHLDAPDSYIMKPVHEDHTIH
VPEHMYILPHKNPOSYVLLMEPKKEINFQOLPPSLFVRLCENLETLQYIPRRHNFLS
Query Match      50.2%; Score 380; DB 8; Length 118986;
Best local Similarity 68.9%; Pred. No. 2.7e-88;
Matches 521; Conservative 0; Mismatches 235; Indels 0; Gaps 0;

```

```

QY 2 aaaaaggttgtaagctctgtatgaaggtatattatcatatgtgtcctaagcattg 61
Db 93547 AAAAGAAATTTTGAACCTGTTGATGCTGTCATCTACCCATCTCTGATGATTG 93606
QY 62 ggtgaagtcggtcgaagcgttccttaagaaggaggaattacggtgtccctaataa 121
Db 93607 GGTTCCTCCAGTCATTCGCTCCCTCAAAAAGATGGAATATGTTGTCAAAATAA 93666
QY 122 ggaatgaatgacccgacagaagatatacactggtcgaagatggtgacgttcagaa 181
Db 93667 AGATGAACATGATCCCTACTAGACTATTAAGTGTGATAGAAATGTCATATTAAGAA 93726
QY 182 actgaataagccactagaagaatacattaccccttgcccttatacgacataatgcta 241
Db 93727 GTTGAATGCTGATCTAGAGAAAGATCATTTTCCTTACCATTCATGACCAATGCTTGA 93786
QY 242 aaggtgtcacaactcacacactctgcttccttcagaagcgttatctagtctcccaat 301
Db 93787 AGCTTTGGCTAATCAATCATTATTTGCTTTCTTGATGATACAGTGGTTCTTTCAAT 93846
QY 302 accagttgcacaactctgacagaagaaccacttcacctgccttcggtacattgc 361
Db 93847 ACCAATTCACCTTAATGATCAAGAAACCACTTTCAGCTGCTTATGAGAACTTTTG 93906
QY 362 ttataagactagccttctgcttgaatgactgcacacttcacacttcgaagatgatat 421
Db 93907 CATTAAAGAAATGCCATTGTTGTTATGCAATGCTCTGCACATTTTCAGAGGTGATGAC 93966
QY 422 ggcattatcttaactctgtgaaataatctcgaagtttcaatgactgttccttccgt 481
Db 93967 CTTATATTTTTCAGACTTAATCAAGAGATGAGTGGAGGTTTTCATGACGATTTTTCGT 94026
QY 482 ttaagggtctcttctgactgactcccaagcacttcgagttctacagagatgtaa 541
Db 94027 CTATGCGCCCTCTCTCTTCATGTTTGTGATCTTGGCAGGATTTGACAGGTGCGA 94086
QY 542 agacccaactctgctcgtgaatgggagagaagtcacacttatagttaagaagcagc 601
Db 94087 AGAGCGAATCTTGTTCATTTGGGAAAAGTGTCAATTTTCATGATGTAAGAGGCAATG 94146
QY 602 cttagagataaataatctgaaagaggtatgaatcgaatgaaggtgaaggtgacat 661
Db 94147 ATTGGGTACAAAGATATCAGAGGAAGGATATGAGTTGACAAAGAAAGTTGAAGTGT 94206
QY 662 cgaacaatgcaatacccacagatatacaaggtataagaagttccttgctgcatgctg 721
Db 94207 GATGACGCTGAGCCCTTAAGAGGTGAGACATCAGAACCTCTCTGTCATGCTGG 94266
QY 722 ttctatagaagttatcaagaattcacaagaagt 757
Db 94267 ATTCTATGAAAGATTCAATTAAGACTTTTCTAAGAT 94302

```

```

RESULT 14
T6L9/c 76L9 54573 bp DNA T6L9 10-JUN-1999
DEFINITION Arabidopsis thaliana BAC T6L9.
ACCESSION AF147265
VERSION AF147265.1 GI:4732170
KEYWORDS

```

SOURCE

ORGANISM

thale cress.

Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.

1 (bases 1 to 54573)

Washington University Genome Sequencing Center.

The A. thaliana Genome Sequencing Project

Unpublished (1997)

2 (bases 1 to 54573)

Cotton, M., Mohlman, P. and Sapetti, L.

The sequence of A. thaliana T6L9

Unpublished (1999)

3 (bases 1 to 54573)

Waterston, R.

Direct Submission

Submitted (30-APR-1999) Department of Genetics, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

4 (bases 1 to 54573)

Waterston, R.

Direct Submission

Submitted (10-JUN-1999) Department of Genetics, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

Submitted by:

Genome Sequencing Center

Department of Genetics, Washington University,

St. Louis, MO 63108, USA

e-mail: twilson@watson.wustl.edu

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone

NEIGHBORING COSMID INFORMATION:

The 5' clone is F10A2, 200 bp overlap; 3' clone is F5K24, 200 bp overlap. Actual start of this clone is at base position 86567 of F10A2; actual end is at 39727 of F5K4.

A tandem repeat extends from base position 33846-49042 in the clone T6L9. Sequence fidelity in this area can not be guaranteed due to possible misassemblies, single clone and single stranded regions, and low consensus quality areas. Digests with EcoRV and HindIII are consistent with the submitted consensus.

NOTES:

Coding sequences below are predicted from computer analysis, using the program GeneFinder (P. Green and L. Hillier, ms in preparation).

```

FEATURES
SOURCE
1..54573
/organism="Arabidopsis thaliana"
/cultivar="Columbia"

```



```

/db_xref="taxon:3702"
/clone="T6L9"
/chromosome="IV"
/map="unknown"
1206..2250
/gene="T6L9.5"
join(1206..1364,1437..1532,1614..1776,2072..2250)
/gene="T6L9.5"
/codon_start=1
/protein_id="A038217.1"
/evidence="not_experimental"
/db_xref="GI:5032269"
/translation="MANEVSILCPDLGIPLYGSDPPTSVLPHYVYEPKPK
KESYDFADKYLEKRRACDEROHLGLRQLEAKIAQEMKMTLGSQGRPDPKRVY
KVPFDESEDEEDINTMKRVKTNPMKPRPQNMEDISTETVDSHADGMI
REVPKPDORPPPLRGVFRSEIWTPTCLIKDPERDO"
complement(3014..3990)
/feature="pseudogene; contains similarity to retroviral
polyproteins"
1394..15837
/gene="T6L9.4"
join(1394..15454,15589..15837)
/gene="T6L9.4"
/feature="similar to Arabidopsis thaliana putative En/Spm
transposon protein (GB:AC005396)"
/codon_start=1
/evidence="not_experimental"
/protein_id="A038214.1"
/db_xref="GI:5032266"
/translation="MDAYKKRCMDYHHGSENSRAESDPTKVYQWNNDELYQAEFLDE
EPAYVIGSEIAECDDKDEFLAKLDAEPLVPCSNHSLSAIYSLFLEPNKNG
SOKSFNDLEIPEMLPKDNYLHSLYKVKFKLSFDMCYOKIHACVNDCLFPRKTK
KUDSCPKNASRMKTNLHIGELKGVOKVLRTPITRLGMRKMDLDMHRS
NKSDBKLEHVDLTWQNNAKYPSEGEERNRHGISTDGPVFNKRNKISCPWD
LCKKENIMTLILIPQOQNNKIDVLEPIEDICHLMDGELTYAASSTEDLKA
MLMTISDPFVAGNLAGCKVKMGCPGCKNTDSMLKFSRKHVYMCHILGLPTTH
FRMKSMFSGKAENGRKRIISGOEYONLKNFNKNEFTLQOSAKKRTVINIADSD
SDOSSESEBBDQVDEBELSRMKKSIFKPLPWELLVRHNDVHVRVNAASI
ISTLHKWKSQDGLNARKDLEVLGRKHRYTKGRITLPAALMSLSKEICTARMS
1KF"
16767..18753
/feature="pseudogene"
complement(22519..27395)
/feature="contains similarity to retrotransposon-like
proteins"
29824..33456
/gene="T6L9.2"
join(29824..29837,29873..30092,30287..30343,30437..30687,
30777..30965,31095..31230,31477..31584,31668..31913,
32000..32038,32118..32417,32580..33119,33205..33456)
/gene="T6L9.2"
/feature="contains similarity to Arabidopsis thaliana
hypothetical protein (GB:AC005396); maybe a pseudogene"
/codon_start=1
/evidence="not_experimental"
/protein_id="A038216.1"
/db_xref="GI:5032268"
/translation="MYLLEPYILVSOAKOVYLRDESSPPTVALRGSSRFSKEE
LDSVAIGLPPNVDIDDLDERENERDCEIYNNAKANEVEITGITEVONDED
HLEPDVANDVDVDDDKVNEEPHGLVLDCEVPRKRRGRTRLDIADK
EOKYRARIADDSQSLVLIIMVNPQVLEFNMGEIDCKSVKLTSLVGAALVRHP
VTEHMTKIGELRTYLVKSVSEFVSMKSRQVTJNREKTDQENMRMKNISFED
RKPVTKTQOEFKRTASADPSFVKRLKVVKRSRTKRDAPVNTNATEKIKKAEIYKS
DNPSSEADDSLSQLGPNPGRILRAMGNNMKTACQCLLOHMGASDAVAAAGCTISS
KNEIDEVIERKMNROSAEVEGNSAARVNRTQPKCTLLHMGASDAVAAAGCTISS
DDELVNDRLGPTDVKFLIDIAIEPEALMRANMFLIEAAGVGMILWPAAMCSL
EEDLOOEDIAQOENALVNGLPGLHVAVKLFVDVAVIPEGFTIMPNTNVYLEBC
MSFVAMPVNVKVIENPADATYOKSPLOTSAARGSPLEATTSGKSAASVSGPSGL
KSAKSKKAVASGAASGOTAGYKPSSEIFPAKGTSAATVSKIAEKSIKAGSKS
PVKSKLDSQLHLRSRPRKGAIEVTKTHKCKIMDISRKOVYAGRVHSDPPQKK
QHVRLGDNAKRVAVDVKIDATVWRASDEIEYMRDSSLSIA"
50747..51653
/gene="T6L9.1"

```

```

CDS
join(50747..50820,51398..51426,51475..51653)
/gene="T6L9.1"
/codon_start=1
/evidence="not_experimental"
/protein_id="A038215.1"
/db_xref="GI:5032267"
/translation="MOTFEVIGYVVLGIPISIKNNMFELCWVKEDGEDEEN
DEDEKEKEEYKNYADGSDNGNENMLTIKIMRLNMKIIORHVLVMI"
BASE COUNT 16547 a 10322 c 9855 g 17849 t
ORIGIN

Query Match 50.0%; Score 378.4; DB 8; Length 54573;
Best Local Similarity 68.8%; Pred. No. 6,4e-88;
Matches 520; Conservative 0; Mismatches 236; Indels 0; Gaps 0;

QY 2 aaaaaggttgtagaagctccctgtagaaggtatlatatcatcgttgctcatgcatg 61
Db 25180 AAAAGAAATTTTGAACGCTTGATGCTGGTGCATCTACCTCATCTGATGACTTGG 25121

QY 62 ggtgagtcggtgtagatgagcttccctaaagaagagagattacggtgctccatgtaa 121
Db 25120 GGTTCCTCCAGTGCATGTCGTCCTTAAGAGGATGAGTACGTCTTCAAAAATGAAA 25061

QY 122 gtagaattgtagcccgagagattatcatctgtagatagatgtagatgtagatgtaga 181
Db 25060 AGATGAACTGATCCCTACTACTAGAACGTCATAGATGATGATGATGATGATGAA 25001

QY 182 actgaataaagccactgagaagaatcatcacccttgctcctttagatgtagatgtaga 241
Db 25000 GTTGAAAGTTCACATCTGGAAGAAAGATCAATTCCTTACCATTCATGACCAATGCTTGA 24941

QY 242 aaggtgtccaaactcaaacactctccttcttagaagattatccttagtctcccaat 301
Db 24940 ACGTTTGCTATATCATCATATTTATGCTTCTTGATGAGATGAGTGGTTCTTCAAT 24881

QY 302 accagttgacaatctgtagcagagaanaaccacttcaactgctccttgtagtaccttgc 361
Db 24880 ACCAATTCACCTATATGATCAAGAAAGAAACACATTCACGTCCTTATGACACTTTTGC 24821

QY 362 ttatagaagtagtcccttgtagttagatgtagcactgtagcacttcaagaagtagatg 421
Db 24820 TTATAGAGAAAGCCATTTGGTTATGCAATGCTCTGCAACATTTCAAGAGTGTATGAC 24761

QY 422 ggcatactccctaaccttgtagaataatgtcgaagtttcttagatgtagtaccttgc 481
Db 24760 CTTTATATTTTATAGCTTAATGACAGAGATGAGTGAAGTTTCAATGACGATTTGCGT 24701

QY 482 ttacggtctctcttgtagatgtagtccctgaacacactgtagtgcagtagttagaagatgaa 541
Db 24700 CTATGGCCCTCTCTCTCTCATGTTGTGTTGAGATGTTGGACAGGATATTGACAGGTGCGA 24641

QY 542 aagcccaatctgtctgtagatgtaggagagagagagagagagagagagagagagagat 601
Db 24640 AAGAGCGAATCTTGTCTCAATTTGGAGAAAGTGTATTTCAATGAGGAGAGAGATGAT 24581

QY 602 cttagaacaataatcttctgaagaagtagatgtagtgcagtagaagcctgaaggttagatgcaat 661
Db 24580 GTTGGGTCAAGATATAGAGAGAGGATATAGAGGTGACAAAGAGAAAGTTGAAGTAT 24521

QY 662 cgacaaaatgtagcattcccaacagatatacaagtagtaagagttccttgtagtgcagtag 721
Db 24520 GATGCGAGTTCAGCCACCAAAAACGGTGAAGGAATTCAGAAATGTTCTTGGTATGCTGG 24461

QY 722 ttctctagaaggttlatcaagaattcccaagaagt 757
Db 24460 GTTCTATAGAGATTATTAAGACTTCTTAAGAT 24425

RESULT 15
LOCUS AB026642/c 77287 bp DNA PLN 27-DEC-2000
DEFINITION Arabidopsis thaliana genomic DNA, chromosome 3, pl clone: MBD5.

```

ACCESSION	AB026642 BA000014	CDS	complement(join(4457, .4848,4900, .5115))
VERSION	AB026642.1 GI:4757398		/note="gb AAD28667.1
KEYWORDS			gene_id:MED5.3
SOURCE	Arabidopsis thaliana (strain:Columbia) DNA, clone_11b:Mitsui P1 clone:MED5.		similar to unknown protein"
ORGANISM	Arabidopsis thaliana		/codon_start=1
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis. 1 (sites)	CDS	/pseudo
AUTHORS	Sato,S., Nakamura,Y., Kaneko,T., Katoh,T., Asamizu,E. and Tabata,S.		/evidence=not_experimental
TITLE	Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence features of the regions of 4,504,864 bp covered by sixty P1 and TAC clones		/product="Athlia retroelement ORF1 protein-like"
JOURNAL	DNA Res. 7 (2), 131-135 (2000)	CDS	complement(join(16691, .17082,17134, .17349))
MEDLINE	20277480		/note="gb AAD28667.1
REFERENCE	2 (bases 1 to 77287)		gene_id:MED5.5
AUTHORS	Sato,S., Nakamura,Y., Kaneko,T., Kato,T., Asamizu,E. and Tabata,S.		similar to unknown protein"
TITLE	Direct Submission		/codon_start=1
JOURNAL	Submitted (28-APR-1999) Yasukazu Nakamura, Kazusa DNA Research Institute, Department of Plant Gene Research: 1532-3, Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:yakamukazusa.or.jp, Tel:81-438-52-3935, Fax:81-438-52-3934)	CDS	/pseudo
COMMENT	Address for correspondence: kaos@kazusa.or.jp For the latest information on annotation of this clone, please see http://www.kazusa.or.jp/kaos/cgi-bin/agd_graph.cgi?c=MED5 Genes with similarity to proteins in the databases are described in 'product' or 'note' qualifiers. Genes that have no significant protein similarity are described as 'unknown protein'. The software programs used to predict genes include: Graal (Informatics Group, Oak Ridge National Laboratory, http://combio.ornl.gov/Graal-1.3/), GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html), NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and SplicePredictor (Volker Brendel, Stanford University, http://gremml1.zool.iastate.edu/cgi-bin/sp.cgi). Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, Washington University School of Medicine, St. Louis, http://genome.wustl.edu/eddy/tRNAscan-SE/). This sequence may not be the entire insert of this clone. It may be shorter because we remove overlaps between neighboring submissions. The 5' clone is MED16 and the 3' clone is F21A17.	CDS	/evidence=not_experimental
FEATURES	Location/Qualifiers		/product="transposon protein-like"
source	1. 77287		complement(join(26389, .27150,27293, .27596,27715, .28429, 28770, .28877,29127, .29173,29504, .29571))
	/organism="Arabidopsis thaliana"	CDS	/note="gene_id:MED5.8"
	/strain="Columbia"		/codon_start=1
	/db_xref="taxon:3702"		/evidence=not_experimental
	/chromosome="3"		/product="Ac-like transposase"
	/clone="MED5"		/protein_id="BAB01787.1"
	/clone_11b="Mitsui P1"		/db_xref="GI:9293884"
CDS	join(689, .829,1056, .1214)		/translation="MKICNOXYIIALGVRRNFIMFGVNGDOEPRRRORTROSGPIN
	/note="gene_id:MED5.1		MEDNVEVRSKTEKMTDNDMEARIKKKKKYKVDVRIAYVTKLAEOTKRVDPNS
	unknown protein"		THMEIEVRSKTEKMTDNDMEARIKKKKKYKVDVRIAYVTKLAEOTKRVDPNS
	/codon_start=1		GNVLAIRWIGGFDLVKMKQIRANKGELHGHFUSKLTQNELIELLAETKRWMLTKRI
	/evidence=not_experimental		KKAKYFSVLDSSPDSSRKQKFTFLRCVDVSTGSPKTEEFFSTFLATKDRGSLKRL
	/protein_id="BAB01785.1"		TLQDALIDIDLKLTIDIRGGYDGNHMMGKHKGVQKSEIYVLDLOPTIGKCRFTMNC
	/db_xref="GI:9293882"		ITLKLSQTRWESRIDSLSKPIRYQASKLIDALFYLAEYSDNPRHSDVESLAEEETGNC
	/translation="MCAWLVFRYFGCWFSSNRGKTEHEVMPNDVFSKNTLSQPRS		IGCFEFLFGMYIWDLLAAVNIYVPIIKRRLVKKKPGGEVERIDESKTKAESFRHH
	HARGRWIQRHCHSIDRKTEPLKSLINDLHWEAFHSPSDANDIEFGGIMT"		YPMNIMDOALINSIEVRFQFOYEKIFPFLVGLDKRKYAAEDDELRTSCMKLEASLKD
	complement(join(2226, .2351,2434, .2661,2796, .3083, 3159, .3326))		VHSDVDGQDALFMELKLTMDVLPKFTTKVPEVLDVFKRKGDCYPTWATYTRILFTIPVS
	/note="gene_id:MED5.2"		IASAERTPPKTLKLNLYLSRYVPOERLNGLALISTEDLVKLDYOKULINEFAGAKAR
	/codon_start=1		RIKMFRIEKKLFCFLFLATFLFLENFLKONKHHILFCERILGH"
	/evidence=not_experimental		complement(36504, .37124)
	/product="transposon protein-like"		/note="gb AAF24530.1
	/protein_id="BAB01786.1"		gene_id:MED5.9
	/db_xref="GI:9293883"		similar to unknown protein"
	/translation="MFTSFFVCWVRFRDGRSLVKTIRIFLOKDPAPYNNSCVP		/codon_start=1
	VDRKRELFLEAKTHMDPLLTGTVOYTFNFIKRLDWDVSTARTTREGPPWIGETL		/evidence=not_experimental
	WGICATWIDTAQKRSRTYSKARLSRNGIGPVHSGPSEIODELPEKIGRPV		/product="Athlia retroelement ORF1 protein-like"
	HUGEVFIETHRSKDSFVDSKSEIKTAQYQONVDRSLALASASAVDSGSRPELT		complement(46183, .46803)
	LDVYTAIFLESTEDSRGNPYGLCLKIDTLAGANRHHSSSSSFOALEERL"		/note="gb AAF24530.1
CDS			gene_id:MED5.11

```
similar to unknown protein"
/codon_start=1
/evidence-not_experimental
/protein_id="BAB01789.1"
/db_xref="GI:9293886"
/translation="MPRIVLHACCFIALLISDELTLSTYLEADLITTNHSTKRSKS
FRISRDSTHSSRGHHHTSPDLBHEVECHLHHQHTTSLHSIMRSASATFARPLL
DLVTPRGVSPSOHLTAYSMTSPRAFFIPHSPTSRKRRRIQLFTRLTRPPSS
TVLNPSTKYEVLSTIRVSGYFAISSMTSPFSQTRSFPSISADLV"
49533..51269
/note="gene_id:MED5.12"
/codon_start=1
/pseudo
/evidence-not_experimental
/product="reverse transcriptase-like protein"
/complement(join(53826..54084,54364..54445,54944..55217,
55368..55565))
/note="gene_id:MED5.13
unknown protein"
/codon_start=1
/evidence-not_experimental
/protein_id="BAB01790.1"
/db_xref="GI:9293887"
/translation="MATTFIHNDGLTITKOSILTRKRNKNOTSHYNHMKPPQKAPOS
POPTFOIDFIDGDFCEEVLSNTOVFTSANKLFRQESLKNPIYIKLKLITPM
KIYKVSQHKHGFNDKTLMLVPTANHOIDEOKIRFQSEEFSAIYVNIIDFAVY
SLSNSTSRLEPFDDVDDETNOFLTWKIALDLDLMTYNHMDSTTSNATFATINLEQ
ETHOCLATIANKTITTFALSGLTIMDNLEVLFIQNDIAQLSLPLPIRDR"
complement(58666..60218)
/note="gb|AA048951.1
gene_id:MED5.14
similar to unknown protein"
/codon_start=1
/pseudo
/evidence-not_experimental
/complement(72081..76294)
/note="gene_id:MED5.15"
/codon_start=1
/pseudo
/evidence-not_experimental
/product="reverse transcriptase-like protein"

BASE COUNT      24106 a 14640 c 14726 g 23815 t
ORIGIN

Query Match      50.0%; Score 378.4; DB 8; Length 77287;
Best Local Similarity 68.8%; Pred. No. 6.7e-88;
Matches 520; Conservative 0; Mismatches 236; Indels 0; Gaps 0;

QY 2 aaaaagaggttggaagctctgtagaaggtatctatcatctggtccatagcagattg 61
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 12489 AAAAGAAATTTGAACTGTTGATGCTGGTGCATCTACCTATCTCGATGATGACTTG 12430

QY 62 ggtgaagtcgcgagtcataagaggaagagcattaccggtgtccccaatgataa 121
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 12429 GGTGTTCCAGTGCATGCTGCTCCCTAAAGAGGTGGAATGACTGTTGTCAAAAAATGAAAA 12370

QY 122 ggaatgaatgatccgcagagatatacactgagctataagatgagatcagatc 181
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 12369 AGATGAATGATCCCTACTAGAACTATACCTGTCATAGATGTCATGATTAAGGAA 12310

QY 182 actgaataaagcacaatgaagaatatacctcttgccctttatcgacacataatg 241
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 12309 GTTGAATGCTGATGAGAAAGATCATTTTCCATTACCATTCATGACCAATGCTTGA 12250

QY 242 aaggtgtccaaactcacacactctgcttctagaacggtattctagttctcccaat 301
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 12249 ACCTTTGGCTAAATATCCATCTACTATGCTTCTTGATGATGATGAGTGTCTTCAAT 12190

QY 302 accaagttgacaatctgataagagaacacacttaccgtcccttcgtaaccttgc 361
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 12189 ACCAATTCACCTTAATGATCAAGAAAAACCACTTTCAGGTGCTCCTTATGAACTTTTGC 12130
```

```
QY 362 ttatagacatgctcttcttgcttatgtaatgcacctgcacacttcaagaatgatat 421
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 12129 CTATAAGAAATGCGCATTTTGTTATTCACACCTCCCGCAACATTCGGAGGTGTATGAC 12070

QY 422 ggcataatctcctaactcttgtaaaaataltgcagagtttcaatgataacttccgt 481
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 12069 CTCTATATTTCAGACTTAATCGAGAGATGTGAGAGTTTCATGAGACATTTTTCGGT 12010

QY 482 ttaaggctctctcttgatgatgtgcccacagcaacttgacgagttcttcaagagatgaa 541
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 12009 CTATGGCCCTCTTCTCTCATGATGTTGTGAAATCTTGGCAGGGTATGACTAGTGCGA 11950

QY 542 agacaccaatcctgtcttgaaatgagagagtgccacttatagtcaatgaaagcatcgt 601
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 11949 AGAATCGAATCTTCTTCTCAATTGCGAAAGTGTGATTCATGATGAGAGAGCATFAG 11890

QY 602 cttagacaataaaattctgaagaaggtaltgaagtcgataaggtcgaatgataat 661
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 11889 ATTGGTGCACAAATATCAGAAAGCGTATAGAGGTTGACAAAGAAAGTGAAGTGAT 11830

QY 662 cgacaataatgccataacccacagatatcaaggtataagaagttccttgatgagtg 721
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 11829 GATGCACCTGTAGCCACCTAAACGCGTGAAGGACATTAGAAAGCTTCTTGATGCTCG 11770

QY 722 ttctataagaagttatacaagaattcacaaagt 757
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 11769 ATTCTATGAAGATTCATTAAGACTTCTCCAAAGAT 11734
```

Search completed: April 1, 2002, 22:42:57
Job time: 9404 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 1, 2002, 20:11:08 ; Search time 159.19 Seconds
(without alignments)
4076.859 Million cell updates/sec

Title: US-09-586-106-62

Perfect score: 757
Sequence: 1 gaagaagaggttgaagctc.....atcaagaattcacaaggt 757

Scoring table: IDENTITY_NUC
Gapop 10.0 , Capext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_1101:*

1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	381.6	50.4	163319	21	AAE22306 Arabidopsis thaliana
2	380	50.2	59590	21	AAE22281 BAC containing rep
3	375.2	49.6	79122	21	AAE22294 BAC containing rep
4	375.2	49.6	94895	21	AAE22302 BAC containing rep
5	375.2	49.6	94895	21	AAE22302 BAC containing rep
6	375.2	49.6	1082138	21	AAE22305 Arabidopsis thaliana
7	371.2	49.0	129021	21	AAE22296 BAC containing rep
8	370.4	48.9	103929	21	AAE22287 BAC containing rep
9	370.4	48.9	134499	21	AAE22286 BAC containing rep
10	366.4	48.4	109973	21	AAE22298 BAC containing rep
11	364.8	48.2	64415	21	AAE22279 BAC containing rep

12	357	47.2	12286	21	AAE35261
13	356	47.0	83390	21	AAE22283
14	356	47.0	90336	21	AAE22289
15	351.8	46.5	92584	21	AAE22288
16	349	46.1	4609	21	AAE35273
17	348.6	46.1	9139	21	AAE35274
18	344.8	45.5	611590	21	AAE22303
19	335.4	44.3	1082138	21	AAE22305
20	333.8	44.1	9829	21	AAE35271
21	333	44.0	40349	21	AAE22278
22	331.4	43.8	12571	21	AAE35272
23	331.4	43.8	64415	21	AAE22279
24	316.6	41.8	96988	21	AAE22290
25	309.4	40.9	600	21	AAE35284
26	302.4	39.9	80450	21	AAE22295
27	273.2	36.1	600	21	AAE35258
28	259	34.2	597	21	AAE35280
29	244.6	32.3	597	21	AAE35286
30	222.6	29.4	109973	21	AAE22298
31	214.6	28.3	10482	21	AAE35275
32	120.2	15.9	90336	21	AAE22289
33	120.2	15.9	121001	21	AAE22284
34	112.4	14.8	163319	21	AAE22306
35	108	14.3	82588	21	AAE22301
36	107.8	14.2	101786	21	AAE22293
37	105.2	13.9	8435	20	AAE83445
38	101.4	13.4	86584	21	AAE22292
39	97.6	12.9	31491	22	AAE35301
40	97.6	12.9	31491	22	AAE22293
41	93.2	12.3	101786	21	AAE22293
42	91	12.0	936	22	AAE58252
43	91	12.0	936	22	AAE58254
44	91	12.0	936	22	AAE58257
45	91	12.0	936	22	AAE58259

ALIGNMENTS

RESULT 1

AAE22306

ID AAE22306 standard; DNA: 163319 BP.

AC AAE22306;

XX 20-MAR-2001 (first entry)

DT Arabidopsis thaliana chromosome 4 centromere.

DE Centromere: mitochondria; vector: ds.

XX Arabidopsis thaliana.

OS Arabidopsis thaliana.

PN WO20005325-A2.

XX 21-SEP-2000.

PD 17-MAR-2000; 2000WO-US07392.

PF 18-MAR-1999; 99US-0125219.

PR 01-APR-1999; 99US-0127409.

PR 18-MAY-1999; 99US-0134770.

PR 13-SEP-1999; 99US-0153584.

PR 17-SEP-1999; 99US-0154603.

PA (UYCH-) UNIV CHICAGO.

PI Preuss D, Copenhagen G, Keith K.

XX WPI; 2000-587529/55.

DR Recombinant DNA construct comprising a plant centromere, useful for

XX producing stably inherited mitochondria which can serve as vectors for

CC	Factors, cytokines, antibodies, and growth factors.
XX	
SQ	Sequence 79122 BP; 23719 A; 15988 C; 16603 G; 22811 T; 1 other:
	Query Match 49.6%; Score 375.2; DB 21; Length 79122;
	Best Local Similarity 68.5%; Pred. No. 4.2e-101;
	Matches 518; Conservative 0; Mismatches 238; Indels 0; Gaps 0;
OY	2 aaaagagctgtggaagccttcgtatgaaagtatactatcatcattgtgcatacgatg 61
Db	67132 aaagaagaatttggaaacttgcttgatgatccattaccatttccatttcgtatgactg 67191
OY	62 ggtgagtcgcggtacgaacgcttcctcaagaaggagcattaccggtgtgcccaatgataa 121
Db	67192 ggttccacagtgcatctgtacccctaanaagcgcagaaatgacggtgtttaaaaatgaaaa 67251
OY	122 ggaatgattgataccgcgcagaggaatatcacctggtcataaagatgtagtgtatcgattcaggaa 181
Db	67252 gaatgattgtalcccacaactaagaacttaacctggtgcataagaatgtgtattgattataga 67311
OY	182 actgaataaagccactaaggaaagatcatiaaccttgccttllatocgacattcgtaga 241
Db	67312 gtcaaattgctcactcagaagaagatcatcttccctacacattcatlaataaatgtlga 67371
OY	242 aaggtgtgccaaactcacacacttcgcttcttcagaaggtattcagtgttccccaaat 301
Db	67372 acgcttagcctaactcacccaactactcgtcttcttgtaggatalagtggttcttccaat 67431
OY	302 accagttgcaccaatctgatacagagaanaaacatttccactgcgcttcttcgtaaccttcg 361
Db	67432 atcaattcacccctaagatcatagaagaaacacactttcaactgaccttatgtgaactttgc 67491
OY	362 ttataagcgtatgcctttggtcgtatgtaatgcacctgcacacttccaagaatgtaatg 421
Db	67492 ttataaagaagtatcatttggttatgtcaatgccccgcgaataattcaagaagtgataaac 67551
OY	422 ggctatatctctcaacttttgtaaatatgtgcagagtttcaatgtgatcatttccgt 481
Db	67552 ctctatatcttcttatttaacgcagagagatggtggaagcttccatgatagtatcttcaag 67611
OY	482 ttaagggctctctcttgatgatactccctcagcaacctgatacgagttcttacagaga 541
Db	67612 ctaatggccctctcttccctccatcgtctgtgaatccttgcgaagttatgtcacaggtgcga 67671
OY	542 agacaccaatctgtcttgaatggggagagctgccacttctatgtttatagaagcactgcg 601
Db	67672 aagacctaactctgtctcaacttgggaagaatgtcatcttcaatcttgatgaaggacatagc 67731
OY	602 cttaggcacataaattcttgaagaagratgtgaagtcataagcgtcaagtgtgtgcatt 661
Db	67732 attgggtacagaagratltagaagaaggatagatgagttgactcgtggaaaaatlttaaagtat 67791
OY	662 cgacaaaatgycataccacacacagatatcaaatgaatagaagtcttccctgtcatgtgtg 721
Db	67792 gatcgagttgcataccacaacacccgtcaagaacalcagaagtctccttgtaacgtcgtg 67851
OY	722 ttctatagaaggttatcacaagaatttacaagaagt 757
Db	67852 gtctccagagatttattaaagttactctccaagaat 67887
RESULT 4	
ID	AAF22302 standard; DNA; 94895 BP.
AC	AAF22302;
XX	
DT	20-MAR-2001 (first entry)
XX	
DE	BAC containing repeats from centromeres 1-4 #25.
XX	
KW	Centromere; michromosome; vector; ds.

XX Arabidopsis thaliana.
 OS
 XX WO200055325-A2.
 PN
 XX 21-SEP-2000.
 PD
 XX 17-MAR-2000; 2000WO-US07392.
 FE
 XX 18-MAR-1999; 99US-0125219.
 PR 01-APR-1999; 99US-0127409.
 PR 18-MAY-1999; 99US-0134770.
 PR 13-SEP-1999; 99US-0153584.
 PR 17-SEP-1999; 99US-0154603.
 XX (UYCH-) UNIV CHICAGO.
 PA
 XX Preuss D, Copenhaver G, Keith K;
 PI
 XX WPI; 2000-587529/55.
 DR
 XX
 XX Recombinant DNA construct comprising a plant centromere, useful for
 PT producing stably inherited microsome which can serve as vectors for
 PT the construction of transgenic plant and animal cells
 PS
 XX Claim 102; Page 804-819; 1449pp; English.
 CC The present invention relates to a recombinant DNA construct of a plant
 CC (Arabidopsis thaliana) centromere. The constructs are useful for
 CC producing stably inherited microsome which can serve as vectors for
 CC the construction of transgenic plant and animal cells expressing
 CC selected proteins such as hormones, enzymes, interleukins, clotting
 CC factors, cytokines, antibodies, and growth factors.
 XX
 SQ Sequence 94895 BP; 28943 A; 18101 C; 18466 G; 29385 T; 0 other;

Query Match 49.6%; Score 375.2; DB 21; Length 94895;
 Best Local Similarity 68.5%; Pred. No. 4.6e-101;
 Matches 518; Conservative 0; Mismatches 238; Indels 0; Gaps 0;

OY 2 aaaaaggtgtgaaagtcctctgaaagatattatctatctatgtgtgtcctcaagcattg 61
 DB 45846 aaaaagaatttgaacagctgtgagtggtgtcctcaacccctcctgataagcttg 45905
 OY 62 ggtgagtcctggtgcatagcgttccctaaaggaagcattacccgtgtccctaaatgata 121
 DB 45906 ggttcttcacagtcattgctccctaaaggggtgaatgactgtgtcctaaatgataa 45965
 OY 122 ggtatgattgacccgacagagatatacactgctatagatgattgattcgaattcaga 181
 DB 45966 ggtatgattgacccctactagaaacgtgactgcatagattgattgactagataagaa 46025
 OY 182 actgaataaagccactaggaagatcattacccctgttcttcttcaacacatagctaga 241
 DB 46026 gctaattgtcatctaggaagatcattccttccattcattcattgtaacagctgga 46085
 OY 242 aaggtgtcccaactcaacacctcgtctcttagaagcgttatctagtttccccaat 301
 DB 46086 aagcttgcataaccacattatgtcttctgattgataagtagttgtcttccaat 46145
 OY 302 accgtgtcacacattgatacagagaaacacattacccgtcccttgcgttaccttgc 361
 DB 46146 accaattcacctaatgatacagagaaacacattcacaatgcttcttgaacatttgc 46205
 OY 362 ttatagacgtatgcttctgtctatgtaagcaccgtccaccccttcaagaattagat 421
 DB 46206 ttataaagaagtcattcgttattgcaatgctccctgcacacatttcaaggggtatgac 46265
 OY 422 ggcatacttctactactttgtgaaatattgtcagagtttcaatgattcattccgt 481
 DB 46266 ctcatacttctcaacttaactcagagagatggtgaggtcttcaagaagcatttttgt 46325

OY 482 ttacgggtctctcttctgattatcctccagcaacattgacagttcttcaagatgtaa 541
 DB 46326 ctatggccctcttcttcccaatgatttctgtgtctgtcagattatgaagcaggtgga 46385
 OY 542 agacaccaattctgtcttgaalggggaagtgccacattatgtatgaatgaagcatt 601
 DB 46386 ggaagcaactctgttcttcaattgggaaagtgctattcattgattgaaagcattag 46445
 OY 602 cttagagacataaattctcgaagaagatgtaagtcgaatgaagcttgaatgacat 661
 DB 46446 gttaggtcacaagatatacgaagaaaggtatagattgacaaaggaatgaagtgat 46505
 OY 662 cgaacaaatgcatatccacacagatatacaaggtataagattcctgtgtcatgtcgg 721
 DB 46506 gatgcagtgtagcagccacagaaatgtgtagagacagaagaagttcctgtgtcatgtcgg 46565
 OY 722 ttctataagaaggttatcaaaagatttcaaaagt 757
 DB 46566 gtctcacagaagattatataaagacttcccaaat 46601

RESULT 5
 AAF22302/C
 ID AAF22302 standard; DNA; 94895 BP.

XX AAF22302;
 AC
 XX 20-MAR-2001 (first entry)
 DT
 XX BAC containing repeats from centromeres 1-4 #25.
 DE
 XX Centromere; microsome; vector; ds.
 KW
 XX

OS Arabidopsis thaliana.
 XX
 PN WO200055325-A2.
 XX
 PD 21-SEP-2000.
 XX
 XX 17-MAR-2000; 2000WO-US07392.
 FE

XX 18-MAR-1999; 99US-0125219.
 PR 01-APR-1999; 99US-0127409.
 PR 18-MAY-1999; 99US-0134770.
 PR 13-SEP-1999; 99US-0153584.
 PR 17-SEP-1999; 99US-0154603.
 XX
 XX (UYCH-) UNIV CHICAGO.

XX Preuss D, Copenhaver G, Keith K;
 PI
 XX WPI; 2000-587529/55.
 DR

XX Recombinant DNA construct comprising a plant centromere, useful for
 PT producing stably inherited microsome which can serve as vectors for
 PT the construction of transgenic plant and animal cells
 PS
 XX Claim 102; Page 804-819; 1449pp; English.

XX The present invention relates to a recombinant DNA construct of a plant
 CC (Arabidopsis thaliana) centromere. The constructs are useful for
 CC producing stably inherited microsome which can serve as vectors for
 CC the construction of transgenic plant and animal cells expressing
 CC selected proteins such as hormones, enzymes, interleukins, clotting
 CC factors, cytokines, antibodies, and growth factors.
 XX

SQ Sequence 94895 BP; 28943 A; 18101 C; 18466 G; 29385 T; 0 other;

Query Match 49.6%; Score 375.2; DB 21; Length 94895;
 Best Local Similarity 68.5%; Pred. No. 4.6e-101;
 Matches 518; Conservative 0; Mismatches 238; Indels 0; Gaps 0;

```

OY 2 aaaaaggtgtgaaagcctcctggaatgaatattatcatatgtgtcatagcattg 61
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 35151 AAAGGATGACTTAACTTATTAGATGTTGGGTTATCTACCCCTTCTCTGTATACACTTGG 35092
OY 62 ggtgagtcggtgcaagcgtctcctcaagaaggagcaatccgtgtcctcaatgataa 121
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 35091 GGTATCTCCAGTTCATCTGCTTCCCTAAAAAAGAGGTATGACAGTGTGTAAAAATTTCTTA 35032
OY 122 ggaatgaattgacccgagagattatcatcgtcgtatgagtggtgcatgcatgaa 181
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 35031 AGATGAACGTATACCCACTAGACTATACCTGACATGAGATGCTATTGACTATAGAA 34972
OY 182 actgaataaagcactagaagaatcattaccccttgccttattacgacatgata 241
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 34971 GTTAATGCTGCTCTAGAAAAAGACATTTCOCATTGCGCTTCATGTGATGATCTAGA 34912
OY 242 aaggtgtccaaactcaacactctgtcttcttagaaggtatcttagttctcccaat 301
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 34911 AAGATTAGCAACCATCTTACTATTGCTTTTGTATGTTATGCGGATTTTTCAAAT 34852
OY 302 accagttgacacatctgatacagagaaacacacacacacacacacacacacacac 361
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 34851 CCTATTTCACCAAAATGACCAAGAAAAACCACTTTCACCTTCTTATGAGACCTTTGC 34792
OY 362 ttataagcgtatgaccttctgtgctatgtatgacacacacacacacacacacacac 421
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 34791 TTACAAGCGTATGCGCTTGGTCTATGTAATGACACACACTTCTTACGCGGTATGAC 34732
OY 422 ggcataatctcctaactcttctgtgaaatattgtcagagtttcaatgagactttcgt 481
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 34731 TTCTATTCTCGGATTTGATAGAGAGATGATGAGATGATGATGATGATGATGATGAT 34672
OY 482 ttacgggtctctcttctgtgagatgctcctcagcaacacacacacacacacacacac 541
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 34671 GTATGCGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 34612
OY 542 agacacacacacacacacacacacacacacacacacacacacacacacacacacac 601
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 34611 AGAGACAAACATGCTGCTGAACCTGGAAGAGAAATGCCATTTTCATGCTTATGAGAAAGCAGTCGT 34552
OY 602 cttaagacataaattctcgaagaaggtatgagtcgataagcgaaggtatgagtcga 661
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 34551 TTTGGGCGACAAATTTCTTAGAGAGGAGATAGAGTTGATTAAGGCTTAAGGTTACGAT 34492
OY 662 cgacaaatgcataccacacacacacacacacacacacacacacacacacacacac 721
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 34491 GATGCACTTACAGCCACCAAAACCTGCAAAAGACATCAGAGATTTCTTGACATGCAAG 34432
OY 722 ttctatagaaggtttatcaaaagtccacaaggt 757
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 34431 ATTTTACAGAGATTCATCAAGATTTCTCCAAAGTT 34396

```

RESULT 6

AAF22305/c

ID AAF22305 standard; DNA: 1082138 BP.

AAF22305;

20-MAR-2001 (first entry)

Arabidopsis thaliana chromosome 4 centromere.

Centromere; microsome; vector; ds.

Arabidopsis thaliana.

WO200055325-A2.

21-SEP-2000.

17-MAR-2000; 2000WO-US07392.

XX

PF

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

```

PR 18-MAR-1999; 99US-0125219.
PR 01-APR-1999; 99US-0127409.
PR 18-MAY-1999; 99US-0134770.
PR 13-SEP-1999; 99US-0153584.
PR 17-SEP-1999; 99US-0154603.
XX (UYCH-) UNIV CHICAGO.
XX Preuss D, Copenhagen G, Keith K;
XX WPI; 2000-587529/55.
XX
XX Recombinant DNA construct comprising a plant centromere, useful for
XX producing stably inherited microsome which can serve as vectors for
XX the construction of transgenic plant and animal cells
XX
XX Claim 68; Page 977-1388; 1449pp; English.
XX
XX The present invention relates to a recombinant DNA construct of a plant
XX (Arabidopsis thaliana) centromere. The constructs are useful for
XX producing stably inherited microsome which can serve as vectors for
XX the construction of transgenic plant and animal cells expressing
XX selected proteins such as hormones, enzymes, interleukins, clotting
XX factors, cytokines, antibodies, and growth factors.
XX
XX Sequence 1082138 BP; 348775 A; 194404 C; 195515 G; 343444 T; 0 other;
XX
XX Query Match 49.6%; Score 375.2; DB 21; Length 1082138;
XX Best Local Similarity 68.5%; Pred. No. 1.3e-100;
XX Matches 518; Conservative 0; Mismatches 238; Indels 0; Gaps 0;

```

```

OY 2 aaaaaggtgtgaaagcctcctggaatgaatattatcatatgtgtcatagcattg 61
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1036291 AAAAGAAATTTTGAACCTGCTGATGTTGGTGTGATCATCTACCCATATCTGATGTA 1036232
OY 62 ggtgagtcggtgcaagcgtctcctcaagaaggagcaatccgtgtcctcaatgataa 121
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1036231 GGTTCCTCCAGTTCATGATGCGTCCCTAAAAAGGGGAGATGATGCTTTCAAAAATGAAA 1036172
OY 122 ggaatgaattgacccgagagattatcatcgtcgtatgagtggtgcatgcatgaa 181
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1036171 GGATGAACCTGATCCCTACTTAACCTGATGCTGCTGATGATGATGATGATGATGAT 1036112
OY 182 actgaataaagcacaatgaagaatcatcacttctgtcctttagaaggtatcttagttctcccaat 241
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1036111 GCTAAATGTTGCATCTAGGAAGAATCATTTCTTACCATTCATGATCAAAATGCTTGA 1036052
OY 242 aaggtgtccaaactcaacacacttctgtcctttagaaggtatcttagttctcccaat 301
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1036051 ACGCTTACCTATACCCCATCTATTTGCTTCTTGATGATGATGATGATGATGATGAT 1035992
OY 302 accagttgacacacacacacacacacacacacacacacacacacacacacacacacacac 361
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1035991 ACCCAATTCACCTTAATGATCAAGAGAAACACTTTCACATGCTCTTATGGAACCTTTGC 1035932
OY 362 ttataagcgtatgaccttctgtgctatgtatgacacacacacacacacacacacacacac 421
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1035931 TTATAAAGAAATGCTATGCTGTTTATGCAATGCTCTCAACATTTGAGAGGTATATAC 1035872
OY 422 ggcataatctcctaactcttctgtgaaatattgtcgaaggttcatgcatgcatgcat 481
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1035871 CTCTATATTTTCAAAATTAACCTGAGAGAGATGATGAGGCTTCAAGAGCATTTTGGT 1035812
OY 482 ttacgggtctctcttctgtatgtatgctcagcaacacacacacacacacacacacacacac 541
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1035811 CTATGCGCCCTCTTTTCTCTCATGTTGTTGATGCTTGCGACAGTTATTCAGAGGTGCA 1035752
OY 542 agacacacacacacacacacacacacacacacacacacacacacacacacacacac 601
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1035751 GGAGACGAATCTTGTCTCAATTTGGGAAAGTGTCTATTTCATGCTGAGAGGAAGCATAGT 1035692
OY 602 cttaagacataaattctcgaagaaggtatgagtcgataaaggtatgagtcgaat 661
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```


XX WPI: 2000-587529/55.
 XX Recombinant DNA construct comprising a plant centromere, useful for
 PT producing stably inherited microchromosomes which can serve as vectors for
 PT the construction of transgenic plant and animal cells
 XX
 PS Claim 102: Page 484-508; 1449pp; English.
 CC The present invention relates to a recombinant DNA construct of a plant
 CC (Arabidopsis thaliana) centromere. The constructs are useful for
 CC producing stably inherited microchromosomes which can serve as vectors for
 CC the construction of transgenic plant and animal cells expressing
 CC selected proteins such as hormones, enzymes, interleukins, clotting
 CC factors, cytokines, antibodies, and growth factors.
 XX
 SQ Sequence 103929 BP; 32987 A; 19310 C; 18547 G; 33085 T; 0 other;

Query Match 48.9%; Score 370.4; DB 21; Length 103929;
 Best Local Similarity 68.1%; Pred. No. 1.3e-99;
 Matches 515; Conservative 0; Mismatches 241; Indels 0; Gaps 0;

OY 2 aaaaagagttgtgaagctccctgagaggtattatctatcatgttgctcgaagattg 61
 DB 18128 AAGGAGATCTTAACTATTAAGTCTGGGTTATCTACCTATCTCGATGACCTTG 18069
 OY 62 ggtgagtcggtgcatagcgttcctcaagaaggagcattacgctgtccctaatgataa 121
 DB 18068 GGTATCTCCAGTTTCACTGGCTTCTTAAAGAGAGTATGACAGTGTAAATCTTAA 18009
 OY 122 ggaatgatccgcagagagattatcaactgctatagatggtgagctatcgaaga 181
 DB 18008 AGATGAACGATACACCTATTAAGTCTGGGTTATCTACCTATCTCGATGACCTTG 17949
 OY 182 acggaataaagccataggaagaatcattaccccttgcttaccatcattgacatgata 241
 DB 17948 GTTAAATGTTGCTCTAGAAAGAACATTTCCCATTTGCCCTTCATTCATATGCTGTA 17889
 OY 242 aaggtgtccaaactacacactctgcttcttagaaggtattctagttctcccaat 301
 DB 17888 AATATTGACAAACCATCTTACTATTTGATTTAGATTTAGCGGATTTTTCCTTAA 17829
 OY 302 accagtgcacactgcatcagaagaagaacattccactgctccttggtacattcgc 361
 DB 17828 CCTTATTCACCCAAATGACCAAGAAACCATTTCTACTTGTCTTATGAGACCTTTC 17769
 OY 362 tttagacgtatgctcttggctatgtatgacactgacacttccaagtatgata 421
 DB 17768 TTACAGAGCTATGCTTTCGTTCTCTGTATGACACCAAGCTTTTACAGGGTATGAC 17709
 OY 422 ggcatactctcaactttgtgaataatctcgaggtttcattgagatccttccgt 481
 DB 17708 TTCTATTTTCGGATTTGATGAGGATGATGAGATTTTCTATGATGATTTCTGT 17649
 OY 482 ttaaggtctcttcttgatgattgctcagcaactgcatcgaagttctacaagaatgaa 541
 DB 17648 GTATGGCTCTTCT 17589
 OY 542 agacacacactctgcttgaatgaggaagttccacttattggttaagaagcatcgt 601
 DB 17588 AGAATCAAACTGCTGTAAGTGGAGAAATGCAATTCATTCATTCATTCATTCATTCAT 17529
 OY 602 cttagagataaattctcgaagaagttatgagttcgaataaggtcgaattgataat 661
 DB 17528 TTTGGGCGCGGAAATTTCTGAGAGGAAATAGAGGTTTGAAGGCTTGAAGTGAAGT 17469
 OY 662 cgacaaaatgcatccacacagatcacaagaagttacgaagttccttgatcgtatg 721
 DB 17468 GATGACGTTACAGCACCACCAAAATGTCMAAGACATCAGAGTCTTCTTGACATGACG 17409
 OY 722 ttctatagaagttatcaagaattcaagaagt 757

DB 17408 ATTTTACAGATATTCATCAGAGTTCCTCAAGTT 17373

RESULT 9

AAFF22286/c
 ID AAF22286 standard; DNA; 134499 BP.

XX AAF22286;

XX 20-MAR-2001 (first entry)

XX BAC containing repeats from centromeres 1-4 #9.

XX Centromere; microsome; vector; ds.

XX Arabidopsis thaliana.

XX WO200055325-A2.

XX 21-SEP-2000.

XX 17-MAR-2000; 2000WO-US07392.

XX 18-MAR-1999; 99US-0125219.

XX 01-APR-1999; 99US-0127409.

XX 18-MAY-1999; 99US-0134770.

XX 13-SEP-1999; 99US-0153584.

XX 17-SEP-1999; 99US-0154603.

XX (UYCH-) UNIV CHICAGO.

XX Preuss D, Copenhaver G, Keith K;

XX WPI: 2000-587529/55.

XX Recombinant DNA construct comprising a plant centromere, useful for
 PT producing stably inherited microchromosomes which can serve as vectors for
 PT the construction of transgenic plant and animal cells
 XX

PS Claim 102: Page 453-484; 1449pp; English.

CC The present invention relates to a recombinant DNA construct of a plant
 CC (Arabidopsis thaliana) centromere. The constructs are useful for
 CC producing stably inherited microchromosomes which can serve as vectors for
 CC the construction of transgenic plant and animal cells expressing
 CC selected proteins such as hormones, enzymes, interleukins, clotting
 CC factors, cytokines, antibodies, and growth factors.

XX Sequence 134499 BP; 41565 A; 25130 C; 25225 G; 42577 T; 2 other;

Query Match 48.9%; Score 370.4; DB 21; Length 134499;
 Best Local Similarity 68.1%; Pred. No. 1.4e-99;
 Matches 515; Conservative 0; Mismatches 241; Indels 0; Gaps 0;

OY 2 aaaaagagttgtgaagctccctgagaggtattatctatcatgttgctcgaagattg 61
 DB 130511 AAGGAGATCTTAACTATTAAGTCTGGGTTATCTACCTATCTCGATGACCTTG 130452
 OY 62 ggtgagtcggtgcatagcgttcctcaagaaggagcattacgctgtccctaatgataa 121
 DB 130451 GGTATCTCCAGTTTCACTGGCTTCTTAAAGAGAGTATGACAGTGTAAATTTCTAA 130392
 OY 122 ggaatgatccgcagagagattatcaactgctatagatggtgagctatcgaaga 181
 DB 130391 AGATGAACGATACACCTATTAAGTCTGGGTTATCTACCTATCTCGATGACCTTG 130332
 OY 182 actgaataaagccatcagaagaatcattaccccttgcttaccatcattgacatgata 241
 DB 130331 GTTAAATGTTGCTCTAGAAAGAACATTTCCCATTTGCCCTTCATTCATATGCTGTA 130272
 OY 242 aaggtgtccaaactacacactctgcttcttagaaggtattctagttctcccaat 301

	CC	The present invention relates to a recombinant DNA construct of a plant
	CC	(Arabidopsis thaliana) centromere. The constructs are useful for
	CC	producing stably inherited mitochondria which can serve as vectors for
	CC	the construction of transgenic plant and animal cells expressing
	CC	selected proteins such as hormones, enzymes, interleukins, clotting
	CC	factors, cytokines, antibodies, and growth factors.
	XX	
SQ		Sequence 109973 BP; 31657 A; 23911 C; 22655 G; 31745 T; 5 other;
	Query Match	48..4%; Score 366.4; DB 21; Length 109973;
	Best Local Similarity	68.0%; Pred. No. 2.1e-98;
	Matches 511; Conservative	0; Mismatches 241; Indels 0; Gaps 0;
OY	3	aagaaggttgtgaagctccttgatgaaggattatatcatcatgtgctatagcgattgg 62
Db	9313	AAGGAATATTATGAAACTGTGGGAAGAAGAAATTATCTACCAATTTCAGATFAGCAGTTGG 93254
OY	63	gttagtcgggtgatgatacgcttccctaagaagggaagattaccgctgtgccctaatgaag 122
Db	93253	GTTAGTGCAGTGTCATGTAATTACTTAAAGAGAGAGATGTTACAGTGTGCAAGATAGAAA 93194
OY	123	gatgaatgatccccgcagaaggaatlacacatggtgctatagatagtgtgcatgcattcagaaa 182
Db	93193	GACGAGCTAATTCCTACTCGACAATCACAGSGTATCATGATGATGATGATTATGAAAG 93134
OY	183	cftgaataaagccaactagaagaagataattacctttgctctttatcgaccatatgctagaa 242
Db	93133	CTGAATGTTGCTAACCAATGAAAGACATTTCCCTTACCATTTATGATGACATGTTGGAG 93074
OY	243	agggtgtcccaactcacacaacttcgcgtcttctagaagglttaagtttcccaanaa 302
Db	93073	AGGTTTAGCAAATCATACGACTATTGCTTCCGTGATGATGATGATGATGATGATGATGATG 93014
OY	303	ccaagtlgacaactcgtatccagagagaacaaccttacctgcgccccttgcgtaccttgct 362
Db	93013	CCGATTTCATCCAGATGACCAGAGAAAAAGCATTTACCTCCCCCTTATGATGATGATGATG 92954
OY	363	tataagcatgatgcctttgtgctatglaatgacotgcaccttcaaaggatgfatgagt 422
Db	92953	TATCGAGAAATGCCCTTGCGCTTTGTATATGCTCTACACATTTCCAGAGAGATGATG 92894
OY	423	gctatatctctaacttttgtgaataatgtgcgaagtttcaatgtaaaccttccgt 482
Db	92893	TCTATCTTTCACAGACATGATGAGGAATATCATGGAAGTTTTCATGATGATGATGATGATG 92834
OY	483	tacgggtcttcttttgatgattgcctgaagaaccttgatcgaagcttaacagagatgtaa 542
Db	92833	TATGATCATGCTTTGAGAGATGCTTGAAGATGCTTACAAAGTGTTGGCAAGATGAG 92774
OY	543	gacaccaatctgtctgtaatgaggagaagltgccactttaigtgttaaagaagcatcgtc 602
Db	92773	GAGAAACATCTAGTTTGAATTTGAAATGGAAGAAATGNCATCTTATGTCGACGATGATGAT 92714
OY	603	ttaagacataaaatttttgaagaagatgtgaagtgtgataaggttaagtttatgcatac 662
Db	92713	CTCGGACACAGAAATTTCTGACATGTTATGAAAGTTGATGAGCTATATATGCAAGTTATG 92654
OY	663	gacaaaaatgcataccccacagatalatcaaaaggtataagaagttcccttgtaagtgt 722
Db	92653	ACAAGTCTTCAAGGCGCTGATATGTTAAAGCATGAGAGGATTTCTTGCAATGCTGGT 92594
OY	723	ttcataagaagtttatcaagaattccacaa 754
Db	92593	TTCTACAGGAGATTCAATCAAAAGACTTTAGCAA 92562
	RESULT 11	
ID	AAF22279	
XX	AAF22279	Standard; DNA; 64415 BP.
XX	AAF22279;	

CC retroelement-containing vectors, cells and plants. Methods are
 CC provided for introducing the retroelements into cells, especially
 CC when the retroelement carries at least 1 agronomically-significant
 CC characteristic (ACS). In a preferred method, a helper cell line
 CC which expresses gag, pol and env sequences is used to enable
 CC transfer of a secondary construct which carries an ASC and has
 CC retroelement sequences that allow for replication and integration.
 CC Claimed isolated nucleic acid molecules comprise a nucleic acid
 CC sequence selected from a retroelement primer binding site,
 CC envelope, gag, integrase, reverse transcriptase, protease or
 CC RNase-H sequence (see A435254-61). Also provided are plant
 CC retroviral particles that are used to transfer the nucleic acids
 CC into plant cells.

XX Sequence 12286 BP; 3748 A; 2540 C; 2767 G; 3231 T; 0 other;

Query Match 47.2%; Score 357; DB 21; Length 12286;
 Best Local Similarity 67.0%; Pred. No. 5e-96;
 Matches 507; Conservative 0; Mismatches 250; Indels 0; Gaps 0;

QY 1 gaaaagaggttgtaagctcttgatgaaggtatctatctatctgtgctcatagcatt 60
 DB 4258 gaagagaggtgactcaagctcttgaggctggtcctatatacccatctcgacagcgtt 4317
 QY 61 ggggtggtccggtgcatagcgttcctaaagaggagcattaccgtgtccctaatgata 120
 DB 4318 gggtaagccagctacaggtgtgtcccaagaaaggtgaatgacagtggtacgagatgaga 4377
 QY 121 agagtaattgatactccgagagatatactgactgctatagatgagtgatcttcaaga 180
 DB 4378 ggaatgactgtataccacaacagcaactgctcactgttgcgcgaatggtatgactatcgca 4437
 QY 181 aactgaataagccactaggaagatcattacccttgccttcttaaccacatctgag 240
 DB 4438 agctgaatgaagccacaggaagacattcccttacccttccatgagatcagatctgag 4497
 QY 241 aaagttgtccaaacactcactctgtcttcttaagagttattcttaagttcccaaa 300
 DB 4498 agagacttgcagggcagagatcactactgttcttctgagtactctgagatacaacaga 4557
 QY 301 taaccgttgcacaatctgacagagaacacattccacggtccgttcgtgactcttg 360
 DB 4558 tcgcgtagaacccacagagtcagaggaagcggccttaccatgaccccttggcgctcttg 4617
 QY 361 ctatagacgtatgctcttctgtctatgtatgacacgtgcacacttcaaaagtatga 420
 DB 4618 ctacagagaagtgatccattcggtgtatgtatgacacacacatttcagaggtgatgc 4677
 QY 421 tggctatactctactacttcttgtaaataatgtctgaagttctctgagatgacttccg 480
 DB 4678 tggccatttttcaacacatggtggaagaaagcactcgaggtattctatgacgacttccg 4737
 QY 481 ttacaggtctcttcttgatgatgtccctgaacacttgcagatcttcaacagatga 540
 DB 4738 ttcttgagccctcatltagacagctgtttgaggaactaactagaggggttacttcgaagtgcg 4797
 QY 541 aagacacaactctgtcttgatggggaaggtgcaccattatggttcaatgaagcatcg 600
 DB 4798 aaggaactaacttggttgaatttgggaaggttcatltagtctcgaggggcatag 4857
 QY 601 tcttagacataaaattctctaaagaggtatgtaagtcgataaggtcgaagttggtgca 660
 DB 4858 tccatagaccacaagatctcaacagagaggtatgaggtgacgagcaagaagatgagcgtca 4917
 QY 661 tcgacaaaatgccaatccacacagatatacaaggtatagaagttccctgtcctatggtg 720
 DB 4918 tcgagaaagctgcacacccactgaaatgttaaggggttagaagttcttcttagagcatgca 4977
 QY 721 gtctctagaaggttatacaaaatctcaaaaggt 757
 DB 4978 gtctctagaaggttatacaaaaggtcttcgaagat 5014

RESULT 13
 AAF22283
 ID AAF22283 standard; DNA; 83390 BP.
 XX
 AC AAF22283;
 XX
 DT 20-MAR-2001 (first entry)
 XX
 DE BAC containing repeats from centromeres 1-4 #6.
 XX
 KM Centromere; chromosome; vector; ds.
 XX
 OS Arabidopsis thaliana.
 XX
 PN W020005325-A2.
 XX
 PD 21-SEP-2000.
 XX
 PF 17-MAR-2000; 2000WO-US07392.
 XX
 PR 18-MAR-1999; 99US-0125219.
 PR 01-APR-1999; 99US-0127409.
 PR 18-MAY-1999; 99US-0134770.
 PR 13-SEP-1999; 99US-0153584.
 PR 17-SEP-1999; 99US-0154603.
 XX
 PA (UYCH-) UNIV CHICAGO.
 XX
 PI Preuss D, Copenhaver G, Keith K;
 XX
 DR WPI: 2000-587529/55.
 XX
 PT Recombinant DNA construct comprising a plant centromere, useful for
 PT producing stably inherited microsome which can serve as vectors for
 PT the construction of transgenic plant and animal cells

XX Claim 102; Page 386-404; 1449pp; English.

CC The present invention relates to a recombinant DNA construct of a plant
 CC (Arabidopsis thaliana) centromere. The constructs are useful for
 CC producing stably inherited microsome which can serve as vectors for
 CC the construction of transgenic plant and animal cells expressing
 CC selected proteins such as hormones, enzymes, interleukins, clotting
 CC factors, cytokines, antibodies, and growth factors.

XX Sequence 83390 BP; 24664 A; 17305 C; 17224 G; 24140 T; 57 other;

Query Match 47.0%; Score 356; DB 21; Length 83390;
 Best Local Similarity 66.9%; Pred. No. 2.3e-95;
 Matches 506; Conservative 0; Mismatches 250; Indels 0; Gaps 0;

QY 2 aaaaagaggttggaagctccctcgatgaaggtattatcatcatggtgctcatagcatg 61
 DB 45441 aagaagatgactactatcttagatgctggtgttcttaccctatctcgatagcactg 45500
 QY 62 ggtgagctccggtgcataggtctcctaagaggagagcattaccgtgttccctaagtata 121
 DB 45501 gttactccagttccggtctcccaaaaaggaggtatgacagttgttaaaaatgctaa 45560
 QY 122 ggtatgaattgaccccgagagatatactacgtgcatagagatgcatgcatgagaa 181
 DB 45561 agatgaactgatacccatagactagacataactcgaatagatggtatgactatagaa 45620
 QY 182 actgataaagccaactagaaagatcatlacccttgcctttatcgacacatgctaga 241
 DB 45621 gttaatgctgctctcttaaaagaacattccocatggcccttattgtaaaatgctaga 45680
 QY 242 aaggttccaactccaacacacttctgttctctgacggttatcttagttctcccaaat 301
 DB 45681 aagattagaacacactccttactatgtctgttgatgagtcatacgagatttttcaaat 45740

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 1, 2002, 20:03:58 ; Search time 1412.44 Seconds
(without alignments)
5759.221 Million cell updates/sec

Title: us-09-586-106-62

Perfect score: 757
Sequence: 1 gaaagaggtgtgaaagctc.....atcaaagattcacaaggt 757

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 11351937 seqs, 537289281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_hic:*
10: gb_est1:*
11: gb_est2:*
12: gb_hic:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rtd:*
20: em_gss_vit:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	303.8	40.1	667	13	AQ956471 LERAK19TR
2	288.4	39.4	680	13	B67757 T24E1ATF TA
3	288.8	38.2	624	13	AQ956470 LERAK19TR
4	266.2	35.2	660	13	AQ967851 LER129TR
5	262.8	34.7	622	13	B26562 T10A4TR TAM
6	259	34.2	720	13	CNS005WS Arabidops
7	252.8	33.4	618	13	AQ966890 LERIM76TR
8	243	32.1	612	13	AQ368125 tox0001F
9	239.8	31.7	746	13	AQ960305 LERPB82TR
10	235.4	31.1	617	13	AQ368305 tox0002G
11	227.8	30.1	482	11	BF657025 OV2_18_CO
12	222.2	29.4	510	13	AQ966889 LERIM76TR

13	221.2	29.2	726	13	AQ958534 LERAV15TR
14	214.8	28.4	510	13	AQ967852 LERV129TR
15	213	28.1	785	13	BH023426 GH_MBB000
16	212	28.0	699	13	BH012654 TDGAM25TH
17	206.8	27.3	694	13	AQ956739 LERAK176TR
18	206.2	27.2	514	13	AQ966888 LERIM76TR
19	205	27.1	629	13	AQ688565 nbdx0077L
20	202.4	26.7	453	13	BH009070 eif3c09.x
21	194.6	25.7	412	13	BH013137 TDGAV07TH
22	194.6	25.7	737	13	BH013137 TDGAV07TH
23	192.8	25.5	645	13	AQ367606 nbdx0042I
24	192.6	25.4	617	13	AQ327960 TDGAD42TH
25	182.2	24.1	649	13	BH011999 TDGED22TH
26	177.8	23.5	524	13	BH143723 TDGED22TH
27	176	23.2	561	13	B27354 T5G10TF TAM
28	175.6	23.2	640	13	BH007538 eef8f06.x
29	171.6	22.7	608	13	AQ365596 nbdx0064I
30	168.8	22.3	412	13	BH009071 eif3c09.x
31	164.4	21.7	885	13	AQ861394 nbdx0016F
32	153.6	20.3	736	13	BH022333 GH_MBB000
33	153.4	20.3	533	13	B27662 T13D9TRB TA
34	149.6	19.8	401	13	B29704 T16A24TRB TA
35	147.6	19.5	499	13	AL095432 Arabidops
36	141.6	18.7	534	13	AL096091 Arabidops
37	140.6	18.6	825	13	B19887 T10A4-SP6.T
38	132.8	17.5	697	13	AL0960306 LEREB82TR
39	132.8	16.9	352	13	AL095992 Arabidops
40	127.8	16.8	489	13	AL088130 Arabidops
41	126.8	16.8	410	13	B27967 T23MTRB TA
42	126	16.6	588	13	AQ273556 nbdx0030D
43	125	16.5	693	13	AQ689636 nbdx0080E
44	124	16.4	1062	13	AL057845 Drosophila
45	121.8	16.1	614	13	A2044562 Gm_UMB001

ALIGNMENTS

RESULT 1	AQ956471	667 bp	DNA	GSS	28-JAN-2000
LOCUS	LERAK19TR	LERAK19TR	LERAK19TR	LERAK19TR	
DEFINITION	Arabidopsis thaliana genomic clone LERAK19, DNA				
ACCESSION	AQ956471				
VERSION	AQ956471.1	GI:6784172			
KEYWORDS	GSS				
SOURCE	thale cress.				
ORGANISM	Arabidopsis thaliana				
REFERENCE	Buell, C.R., Lin, X., Pal, G., Barnstead, M., Bowman, C., Ullrich, T.,				
AUTHORS	Feildlyum, T., Liang, F., Creasy, T. and Fraser, C.M.				
TITLE	Genomic survey sequencing of Landsberg erecta ecotype of Arabidopsis thaliana and identification of sequence-based polymorphisms				
JOURNAL	Unpublished (2000)				
COMMENT	Contact: Xiaoying Lin The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0200 Fax: 301 838 0208 Email: atc@igf.org For additional information, see http://www.tigr.org/tdb/at/at.html Seq primer: TR Class: Shocgun.				

FEATURES

source
Location/Qualifiers
1..667
/organism="Arabidopsis thaliana"
/strain="Landsberg erecta"
/db_xref="taxon:3702"
/clone="LERAK19"

/clone_11d="LERA"
 /note="Organ: Leaf; Vector: pHOSt; Total genomic DNA was
 sheared to 0.9-1 kbp before ligation."
 BASE COUNT 203 a 111 c 160 g 192 t 1 others
 ORIGIN

Query Match 40.1%; Score 303.8; DB 13; Length 667;
 Best Local Similarity 69.3%; Pred. No. 7.6e-73;
 Matches 413; Conservative 0; Mismatches 183; Indels 0; Gaps 0;

QY 3 aaagaggtgtgaagctccctgagatgaagatatactatcatctgtgcatagagattg 62
 DB 71 AAAAGAGATTATGAAACCTTGGAAGAGGAGATATCTATCCAAATGTTAGATGAGCATTAG 130
 QY 63 gtggccgggtgcatagcttcctaaagaagaagcattaccgtgtcccaatgataag 122
 DB 131 GTTAGTCACTTCTATGTTGGTTCCTTAAGAGAGGAGTGTACAGTATCAAGAAATGAGAAA 190
 QY 123 gatgaattgataccgcagagatatactacgtgcatagagtgatgagatcagagaa 182
 DB 191 GAGAGAGCTGATTCCTACTCGACAAATCATAGACATGCGATGTCATCGATTACAGAAAG 250
 QY 183 ctgaataaagccaactagagaagatcattacccttgcctttatcgacatagatgaa 242
 DB 251 CTGAATGCTCTACAGAGAAAGACCATTTCCCTTACCATTTATCGATCAGATGTTGAG 310
 QY 243 aggtgtccaaactcaacacctctgtcttagagaggttatctatcttcccaata 302
 DB 311 AGGTTAGCAATATATAGTACTATGCTCTCTTGATGATGATCTCAATGATCTTTCGATC 370
 QY 303 cgaattgcacatctgatacagagaagaaccattcaacctgccttccgttaccctgtc 362
 DB 371 CCCAATCATCCAAATGACCGAGAGAAACGACTTTCACCTGACCCCTATGATGATGCT 430
 QY 363 tatagagctatgctcttctgtatgaatgaacccgtccaccccttcaagaatgata 422
 DB 431 TATCAAGATGCTCCCTGCTCTTGTATGATGCTCTGCGACATTTTCAGAGAGATGATG 490
 QY 423 gctatatctcaactctgtgagaataatgtcgaggttctcagtgatgaccccttccgt 482
 DB 491 TCTATGTTCAAGATATGATCGAGATATCATGATGAGGTTTTCATGATGCTTTCAGTT 550
 QY 483 taaggtctcttctgtatgattgctcccaaaccttgatcagatcttcaagaatgata 542
 DB 551 TATGATCATCATGATGAGATGTTTGAAGAAATCTCTGCAAAAGTGTATGATGATGAG 610
 QY 543 gacaccaatctgtcttgaatgagagaagtgccacttctatgattgaatgaagcat 598
 DB 611 GAGAAACATCTATATTTGATGATGGGAGAAATGTCACTTATGATGATGATGAGAT 666

RESULT 2

B67757 680 bp DNA GSS 09-DEC-1997
 LOCUS T24F14TF TAMU Arabidopsis thaliana genomic clone T24F14, DNA
 DEFINITION sequence.
 ACCESSION B67757
 VERSION B67757.1 GI:266511
 KEYWORDS GSS.
 SOURCE
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 REFERENCE 1 (bases 1 to 680)
 Rounsley, S.D., Field, C.E., Bass, S., Linher, K., Linher, K., Golden, K.,
 Berry, K., Granger, D., Sub, E., Wible, C., Adams, M.D. and Venter,
 J.C.
 TITLE A BAC End Sequence Database for Identifying Minimal Overlaps in
 Arabidopsis Genomic Sequencing. Update 3
 JOURNAL Unpublished (1997)
 COMMENT Other_GSSs: T24F14TR

Contact: Steve Rounsley
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: rounsley@igf.org
 Seq primer: M13-21
 Class: BAC ends
 High quality sequence stop: 680.

FEATURES

source
 1. 680
 /organism="Arabidopsis thaliana"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /clone="T24F14"
 /clone_11d="TAMU"
 /sex="hermaphrodite"
 /note="Vector: BelOBAclI; Site_1: HindIII; Site_2: HindIII
 ; Produced by Rod Wing"
 BASE COUNT 192 a 143 c 144 g 201 t
 ORIGIN

Query Match 39.4%; Score 298.4; DB 13; Length 680;
 Best Local Similarity 70.1%; Pred. No. 2.3e-71;
 Matches 401; Conservative 0; Mismatches 171; Indels 0; Gaps 0;

QY 186 aataagccactagagaagatcattacccttgccttcttaccacatagcagaaag 245
 DB 1 AATGACGCCCATAGAGAAATCATTTCCCTTACCTTCAATGACCAATGCTAAGAAAGA 60
 QY 246 ttgtcacaactcacacactctgtctcttagaaggtattcttagttctcccaataca 305
 DB 61 TTAGCCATATCACCCCTATCTGTTCTTAGATGAGGATATGCGGATTTTTCAAATCCCT 120
 QY 306 gtgcacaatctgatacagagaagaaccattcaacctgccttccgttacccttctat 365
 DB 121 ATCCACCCATATATATCAAGAAATATCAATTCACGTCGCTATGAGGACTTTTGTCTTAC 180
 QY 366 agacgtatgctcttctgtatgtaatgacacctgcctgcacaccttcaagaatgata 425
 DB 181 AAACGATGCTCTTGTGGCTTTGTATGATGACCCACCCACTTTTGAAGCGCTGATGACTTCT 240
 QY 426 atattcctcaactctgtgagaataatgtcgaggttctcagtgatgaccccttccgt 485
 DB 241 ATTTTCTCCGATTTGATAGGAGATGTCGAGGATATCTATGATGATATTTTCTGTAT 300
 QY 486 gggctctcttctgtatgattgctcccaaaccttgatcagagcttctacagatgtaa 545
 DB 301 GGCTCTTCTTCTCTCTCTGTTTGTCAATTTTTCAGAGGCTACTCAAGATGCGAAGAA 360
 QY 546 accaatctgtcttgaatgagagaagtgccacttctatgattgaagaagcagctctta 605
 DB 361 ACAAACTTATGTCGTAACGCGGAAATGTCACTTATGATGATGATGATGATGATGATG 420
 QY 606 gacataaactcttgaagaagtgatgagatgagaaagtgaaagttatgataatcagac 665
 DB 421 GGGCACAAAATATCAAGAAATGAGATGAGATGAGATGAGATGAGATGAGATGAGATG 480
 QY 666 aaaaatgcatacccaagagatatacaaggtataagaagttctctgtcatggtgttc 725
 DB 481 CAGTTGCAACACCAAAATCTGTAAGATATTTAGAAAGTTTCTGCGACATGCTGGTTC 540
 QY 726 tatagaaggttataagaattcccaagatg 757
 DB 541 TACAGGAGATCATCATGATATTTCTCTAGATT 572

RESULT 3
 A0956470/c 624 bp DNA GSS 28-JAN-2000
 LOCUS A0956470/c
 DEFINITION LERAK19TF LERA Arabidopsis thaliana genomic clone LERAK19, DNA

Accession	Version	Keywords	Source	Organism	Reference Authors	Title	Journal Comment	Features	Base Count	Origin
sequence.										
Accession	Version	Keywords	Source	Organism	Reference Authors	Title	Journal Comment	Features	Base Count	Origin
A0956470	1	GI:6784171								
sequence.										
A0956470	1	GI:6784171								
sequence.										
A0956470	1	GI:6784171								
sequence.										
A0956470	1	GI:6784171								
sequence.										
A0956470	1	GI:6784171								
sequence.										
A0956470	1	GI:6784171								
sequence.										
A0956470	1	GI:6784171								
sequence.										
A0956470	1	GI:6784171								
sequence.										
A0956470	1	GI:6784171								
sequence.										
A0956470	1	GI:6784171								
sequence.										
A0956470	1	GI:6784171								
sequence.										
A0956470	1	GI:6784171								
sequence.										
A0956470	1	GI:6784171								
sequence.										
A0956470	1	GI:6784171								
sequence.										
A0956470	1	GI:6784171								
sequence.										
A0956470	1	GI:6784171								
sequence.										
A0956470	1	GI:6784171								
sequence.										
A0956470	1	GI:6784171								
sequence.										
A0956470	1	GI:6784171								
sequence.										
A0956470	1	GI:6784171								
sequence.										
A0956470	1	GI:6784171								
sequence.										
A0956470	1	GI:6784171								
sequence.										
A0956470	1	GI:6784171								
sequence.										
A0956470	1	GI:6784171								
sequence.										
A0956470	1	GI:6784171								
sequence.										
A0956470	1	GI:6784171								
sequence.										
A0956470	1	GI:6784171								

QY	667	aaagccatacccccagatatacaagatataagaagatttccttgctacgtggtcttc	726
Db	144	GTCTTCAAGCGCCGATATATGTCTCAACACGACTGAGAGCTTCTTGACATGTTGGTTCT	85
QY	727	atagaagttatcaagaattca	750
Db	84	ATAGAGATTTCATCAAGACTTTA	61
RESULT	4		
LOCUS	A0967851/c		
DEFINITION	LERIV29TF LERG Arabidopsis thaliana genomic clone LERIV29, DNA		
ACCESSION	A0967851	660 bp	DNA
VERSION	A0967851		
KEYWORDS	A0967851.1	GI:6795552	
SOURCE	GSS.		
ORGANISM	thale cress.		
REFERENCE	Arabidopsis thaliana		
AUTHORS	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.		
TITLE	1 (bases 1 to 660) Buell,C.R., Lin,X., Pat,G., Barnstead,M., Bowman,C., Uterbach,T., Feldblyum,T., Liang,F., Creasy,T. and Fraser,C.M. Genomic survey sequencing of Landsberg erecta ecotype of Arabidopsis thaliana and identification of sequence-based polymorphisms		
JOURNAL	Unpublished (2000)		
COMMENT	Contact: Xiaoying Lin The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0200 Fax: 301 838 0208 Email: atetigr.org For additional information, see http://www.tigr.org/tdb/at.html Seq primer: TF Class: Shotgun.		
FEATURES			
source	Location/Qualifiers 1..660 /organism="Arabidopsis thaliana" /strain="Landsberg erecta" /db_xref="taxon:3702" /clone="LERIV29" /clone="LERIV29" /note="Organ: Leaf; Vector: pUC19UK; Total genomic DNA was sheared to 0.4-0.7 kbp before ligation." BAC COUNT 195 a 127 c 134 g 204 t		
ORIGIN			
Query Match	35.2%;	Score 266.2;	DB 13; Length 660;
Best local similarity	67.7%;	Pred. No. 1.6e-62;	
Matches 373; Conservative	0;	Mismatches 178;	Indels 0; Gaps 0.
QY	2	aaaagagttgtgaagctcctcgatgaaggtatcatcatcagtcgtgcataagcagatg	61
Db	555	AAAGAGATACTTAAATATTATGATGGGGGTGTAATCTATTCACATATTCATGACACTTG	496
QY	62	ggtagcccgatgcatagcgtccctaaagaagggagcattaccgtgtgccctaaagtata	121
Db	495	GGTATCTGTCAGTTCATTTGTGTCTAAAGGGGAGGTATGACAGTTATTAATAATTCAA	436
QY	122	ggatgaattgattcccgacagagatatacaactgctgctatagaatggatgcatcagaa	181
Db	435	AGATGAATGTGTATCCACTATGGACCATACAGGACATAGATGTGTAATTACTATATGGA	376
QY	182	actgaataaagccactagaagaagatcatatcccttgcctttatcgaccataatgctaga	241
Db	375	GCTTATATGACACCACTGAGAAAGATCACTTCCCTTACCTTTCATTTGACCAAAATGCTAGA	316
QY	242	aagttgtccaactacacacactcttgctttctatagaaggtattatcagtttccccaat	301

Db 315 AAGATTAGCAATCACCCCTATTACTGTTTCTAGATGGCTATAGCGGATTTTCAAAAT 256
 Oy 302 accagttgcaaatcgtatagagaagaacacacttcaactgaccttgcgtacacttgc 361
 Db- 255 CCTATCCACCAACCAAGACCAAGAAACTACATTACAGTGCCTATGGGACTTTGCG 196
 Oy 362 ttatagacgtatgaccttgcgtatgtaactgcaactgcaacttcaagaatgtagt 421
 Db 195 TTATTAACGAATGCTTTGGGCTTTGTATGACCCGCCACTTTTCAAGCGCTGAC 136
 Oy 422 ggtatatacttcaacttctgtgtaaaatctgcagagtttcaactgacacttccgt 481
 Db 135 TTTCTATTGTCGATTTGATAGAGAGATGTCAGATTCATCATGATTTTCTGT 76
 Oy 482 ttacaggtctcttctgtgacgtcctcagcaacttgcagcttgcagcttcaagaatgta 541
 Db 75 GTATACCTCTTCT 16
 Oy 542 agacaccaatc 552
 Db 15 AAGACCAAC 5

RESULT 5
 LOCUS B26562 622 bp DNA GSS 13-OCT-1997
 DEFINITION T10A4TR TAMU Arabidopsis thaliana genomic clone T10A4, DNA
 ACCESSION B26562
 VERSION B26562.1 GI:2512528
 KEYWORDS GSS.
 SOURCE thale cress.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 1 (bases 1 to 622)
 Rounsley, S.D., Kelley, J.M., Field, C.E., Craven, M.B., Adams, M.D. and
 Venter, J.C.
 Use of a BAC End Sequence Database To Identify Minimal Overlaps for
 Arabidopsis Genomic Sequencing
 Unpublished (1997)
 Other_GSSs: T10A4TF
 Contact: Steve Rounsley
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0208
 Fax: 301 838 0208
 Email: rounsley@tigr.org
 Seq primer: M13 Reverse
 Class: BAC ends
 High quality sequence stop: 622.
 Location/Qualifiers
 1..622
 /organism="Arabidopsis thaliana"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /clone="T10A4"
 /clone_1lb="TAMU"
 /sex="hermaphrodite"
 /note="Vector: BelobACII; site_1: HindIII; site_2: HindIII
 ; Produced by Rod Wing"

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

BASE COUNT 179 a 128 c 132 g 183 t
 ORIGIN

Query Match 34.7%; Score 262.8; DB 13; Length 622;
 Best Local Similarity 68.2%; Pred. No. 1,4e-61;
 Matches 394; Conservative 0; Mismatches 182; Indels 2; Gaps 2;
 Oy 180 aaactgataagccactaggaagatacattacccttgccttattgcacatatgcta 239

Db 2 AATCTTAATGACGCGCTTGAAGAAAGATCATTTCTATTATCTTATGATCAATGCTA 61
 Oy 240 gaaaggttgcacaactcaacactccttgccttctagaaggtattcttagttccaa 299
 Db 62 GAAAGATTAGCAAAACCATCCCTACTACTGTTTCTAGATGATGATGATGATGATGAT 121
 Oy 300 ataccagttgcaaatcgtatagagaagaacacacttcaactgaccttgcgtacactt 359
 Db 122 ATTCTATCCACCCCAAGACGATCAAGAAACTACTTCTCTCTCTCTCTCTCTCTCT 181
 Oy 360 gcttataaggtatgaccttgcgtatgtaactgcaactgcaacttcaagaatgtagt 419
 Db 182 GCTTACAAGCGAATGCCCTTTGGGCTTTGGCAACGACCTGCTACTTTTCAAGCGGTATG 241
 Oy 420 atgctatattctccttcaacttctgtgaaatatttgcagagtttcaactgacacttcc 479
 Db 242 ACTTCTATTCTCTCGATTTGATAGAGAGACGATGATGATGATGATGATGATGATGAT 301
 Oy 480 gttacaggttcttcttgcgtatgtaactgcaactgcaactgcaactgcaactgca 539
 Db 302 GTGATGATGCT 360
 Oy 540 aaagacaccaatctgtcttgcgtatgtaaggtggaagtgcacatttgcgtatgtaagga 599
 Db 361 CAGGAACCAATCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
 Oy 600 gttctagacataaattcttgcgaagaggtatggaagtgcgaactgcaactgcaactg 659
 Db 421 GTCTTGGGCGACAGAAATCTCAGAAAGGATGATGATGATGATGATGATGATGATGAT 480
 Oy 660 atgcacaaatgcatatccccacagatatcaagaagtcttcttgcgtatgtagt 719
 Db 481 ATGATGCAATTAACGCCCAACAGACTGTGAAAGACATCAAGATTTCTTGGCATCA 540
 Oy 720 gttctatagaaggttattcaagaattccacaaggt 757
 Db 541 GCTTT-TATAGAGATTTCATCAAGACTTCTCCAACT 577

RESULT 6
 LOCUS CNS00SWS 720 bp DNA GSS 28-JUN-1999
 DEFINITION Arabidopsis thaliana genome survey sequence T7 end of BAC T2P6 of
 TAMU library from strain Columbia of Arabidopsis thaliana, genomic
 survey sequence.
 ACCESSION AL088954
 VERSION AL088954.1 GI:5290094
 KEYWORDS GSS.
 SOURCE thale cress.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 1 (bases 1 to 720)
 Salanoubat, M., Cholsen, N., Artiguenave, F., Brothier, P., Wincker, P.,
 Samson, D., Saurin, W., Weissenbach, J. and Queller, F.
 Unpublished
 2 (bases 1 to 720)
 Genoscope.
 Direct Submission
 Submitted (25-JUN-1999) Genoscope - Centre National de Sequencage;
 BP 191 91006 EVRY cedex - FRANCE (E-mail: seqref@genoscope.cns.fr
 - Web : www.genoscope.cns.fr)
 Location/Qualifiers
 1..720
 /organism="Arabidopsis thaliana"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /clone_1lb="TAMU"
 /clone="T2P6"
 /note="end : T7"

FEATURES
 source

BASE COUNT 191 a 156 c 173 g 200 t

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum; Lycopersicon.

REFERENCE 1 (bases 1 to 612)

AUTHORS Budiman, M.A., Mao, L., Wood, T.C. and Wang, R.A.

TITLE A deep-coverage tomato BAC library and prospects toward development

JOURNAL of an STC framework for genome sequencing

MEDLINE Genome Res. 10 (1), 129-136 (2000)

COMMENT

Contact: Wang RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rtwing@clemson.edu
Seq primer: GGAACACGCTATGACCATG
Class: BAC ends
High quality sequence stop: 470.

FEATURES

Source

1. 612

/organism="Lycopersicon esculentum"

/cultivar="Heinz 1706"

/db_xref="taxon:4081"

/clone="tox0001f24r"

/clone_lib="CUGI Tomato BAC Library"

/tissue_type="Nuclei preparation from leaf"

/lab_host="E. coli DH10B"

/note="Vector: pBelobAC 11; Site 1: HindIII; Site 2: HindIII; Tomato is a vegetable crop that ranks second only to potatoes in value and importance. Among plant geneticists and physiologists, tomato represents an ideal dicot model beside Arabidopsis and monocot rice to derive genomic information from. To facilitate the genome analysis of tomato, we have constructed a tomato BAC library that is suitable for positional cloning, physical mapping, and genome sequencing. The library contains 129,000 clones and a random sampling of 498 clones indicated an average insert size of 117.5 kb. With 15X haploid genome equivalents (1C equals 953 Mb) (Arumuganathan and Earle, 1991), the probability to recover any particular sequence is greater than 99%. High stability, large insert and ease in manipulation make BAC libraries the choice for genome sequencing. Pre-characterization of a few hundred bases of insert ends will make BAC clones extremely useful for rapid contig assembly (Venter, Smith, and Hood, 1996). Here we present the construction, characterization of the tomato BAC library, and preliminary analysis of the 1536 tomato BAC end sequences."

BASE COUNT 192 a 149 c 88 g 182 t 1 others
ORIGIN

Query Match 32.1%; Score 243; DB 13; Length 612;
Best Local Similarity 62.4%; Pred. No. 3,7e-56;
Matches 381; Conservative 0; Mismatches 220; Indels 0; Gaps 0;

Oy 90 aaggagcattaccgtgtccctcaatgaataagatgaattccgcagaggaattatc 149
Db 611 AAAGGGGAATGATTGATGCCCAATGAGAAATGAACATTTGTCATGACAGAACGTT 552
Oy 150 actgataagatgagatgatttcagaaactgaataagcactgaagaatcat 209
Db 551 ACTGATGAGAGGTGTGATGATTTACACAAATTAATGATGATGATGAAGAACAT 492
Oy 210 tacccttgcttaccatagatagatgaagaagtgtccaaactcacacactctgc 269
Db 491 TTTCCTATGCCCTCATGATGATTAATGATGATGATGATGATGATGATGATGAT 432
Oy 270 ttttaagagttattctagttctcccaataacagatgacaaatctgacagagaa 329
Db 431 TTTCCTATGATGATTTGGGTATATATCAATGATTTTATTCACACAGAACATCAAGAGAA 372

Oy 330 accacttaccctgccttcgtgaccttgccttataagacgataccttggcctatgt 389
Db 371 ACCACTTACTTTCATATGAGACCTTTCGCTTCAAAAAATGTTGGTGGTTC 312
Oy 390 aatgcacccgacacttaagaatgataagatgatacttctacttctgtaaat 449
Db 311 AATGACCCGCCACATTTACATGATGATGATGATGATGATGATGATGATGAT 252
Oy 450 attgcagagtttcaatgataccttccgttaccggtctcttctgatactgctc 509
Db 251 ATTATAGAGCTTTTATGATGATTTTGGTGGTGGTGGTGGTGGTGGTGGT 192
Oy 510 agcaacatgatacgttcaagagatgataagacccaactctgtctgaatggag 569
Db 191 AACCAATTTATCTGAGGCTCTTATGAGATGTAAGACTGCATTTAGTAAATGGGGA 132
Oy 570 aagtcacacttaltgataagagcactgccttaagacataaattctgaaagagt 629
Db 131 AATGCCACTTCATGATGTAAGAAATGATTTGTTGGTCAATCCGATTTCAAAAAAGGCG 72
Oy 630 attgaagtcgataagagctgaatgatacgaacaaatgcataccacagatattc 689
Db 71 ATAGAGTTGATGAGCTAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGT 12
Oy 690 aaggatgaag 700
Db 11 AAAGGTGTGAG 1

RESULT

LOCUS

AO960305 746 bp DNA GSS 28-JAN-2000
DEFINITION LRRFB82TF LERA Arabidopsis thaliana genomic clone LRRFB82, DNA

ACCESSION

AO960305
VERSION AO960305.1 GI:6788006

KEYWORDS

SOURCE

ORGANISM

Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 746)
Buell, C.R., Lin, X., Pal, G., Barnstead, M., Bowman, C., Utterbach, T.,
Feldblum, T., Liang, F., Creasy, T. and Fraser, C.M.
Genomic survey sequencing of Landsberg erecta ecotype of
Arabidopsis thaliana and identification of sequence-based
polymorphisms
Unpublished (2000)

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Contact: Xiaoying Lin
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: atc@tigr.org
For additional information, see <http://www.tigr.org/tdb/at/at.html>
Seq primer: TF
Class: shotgun.

FEATURES

source

1. 746
/organism="Arabidopsis thaliana"
/strain="Landsberg erecta"
/db_xref="taxon:3702"
/clone="LRRFB82"
/clone_lib="LERA"
/note="Organ: Leaf; Vector: PHOS1; Total genomic DNA was sheared to 0.9-1 kbp before ligation."
BASE COUNT 220 a 123 c 175 g 228 t
ORIGIN

Query Match 31.7%; Score 239.8; DB 13; Length 746;
Best Local Similarity 66.3%; Pred. No. 2.9e-55;

	Matches	360:	Conservative	0:	Mismatches	182;	Indels	1:	Gaps	1,
Oy	213	ccttgcctttatcagaccatacgtagaaaggttgcccaactcacacacttcgctt	272							
Db	9	CCCCCTTCATTTTATTCAGAACCAATGTGGAGGATTGACGAATAATCGAAACTACTACTGTTTT	68							
Oy	273	ctgaagcgtaattctagtcttcgccaaataccagttgcacatcgtatcagagaanaaac	332							
Db	69	CTTAGTGATATTTACAGGTTTCTTCCAAATGCCGATTCATCTCGAATGACCAGAGAAAAT	128							
Oy	333	aacttacctgcoccttcgylacacttggcttatagaaglatgaccttggcttatgtaac	392							
Db	129	ACTTTCACCTTGCACAATATGTATCATTTGCTTATCGAAMATGCCYTTGCGTCTTGTGAAC	188							
Oy	393	gcacctgc-caaccttsaagaatgatagtgatatactctctaacttggtaaataat	451							
Db	189	GCTTCCGAAACCTTTCAGAGAGGTATATGCTTATCTTCACGATATGATTTGAGAGATAT	248							
Oy	452	tgtcgaaggtttcatagtatgaacttccgcttacggtctctctttgatlgaattgcctcag	511							
Db	249	CATGGAAGGCTTTATTTGATGATTTTTCAGTCTATGAGCACCATCATATGTAGAGACTGTTTGA	308							
Oy	512	caacctltatcgaagctctacagagatgtgaaaagaccacatcttgccttgaatbgggagaa	571							
Db	309	GAATCTTCGCCAAGATTTGTTGGAGATGTGAGAGAAACATCTAGTTTGAAATGGGAAAAA	368							
Oy	572	gtgcacacttcatgtgttaatgaaagcatgcctcttaagacataaaacttgaagaagqat	631							
Db	369	ATGTACTTATTAGTGCTAGATGGATGATGTTCTTGTCACATGATTTCTGAGCANTGAT	428							
Oy	632	tgaagtcgataagcgtlaaggttgtatgatactgcacaataatgcccataccaagatacaa	691							
Db	439	AGAGGTTGTATAGCAAGATGCAAGATGCAAGTCACAAACGCTTCAGGGCCCTGGAAAGTGTAA	488							
Oy	692	aggtaataagaattccctctggtcatcggatggtttctcatagaaggttatacaaatctcac	751							
Db	489	AGCAGTGAAGAGATTTCCTAGTCATGCAATGTTTTCACAAACATTCATCAAGAATTTTAA	548							
Oy	752	aaa 754								
Db	549	AAA 551								
RESULT	10									
AQ368305/c										
LOCUS	AQ368305	617 bp	DNA	GSS				07-MAR-2000		
DEFINITION	tox00002G17r CUG1 Tomato BAC Library Lycopersicon esculentum genomic clone tox00002G17r, DNA sequence.									
ACCESSION	AQ368305									
VERSION	AQ368305.1 GI:4221582									
KEYWORDS	GSS.									
SOURCE	tomato.									
ORGANISM	Lycopersicon esculentum									
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum; Lycopersicon.									
REFERENCE	1 (bases 1 to 617)									
AUTHORS	Budiman,M.A., Mao,L., Wood,T.C. and Wing,R.A.									
TITLE	A deep-coverage tomato BAC library and prospects toward development of an STC framework for genome sequencing									
JOURNAL	Genome Res. 10 (1), 129-136 (2000)									
MEDLINE	20113122									
COMMENT	Contact: Wing RA Clemson University Genomics Institute 100 Jordan Hall, Clemson, SC 29634, USA Tel: 864 656 7288 Fax: 864 656 4293 Email: rwing@clemson.edu Seq primer: GGAAACAGCTATGACCATG Class: BAC ends High quality sequence stop: 481.									

FEATURES	source	Location/Qualifiers
		1..617
		/organism="Lycopersicon esculentum"
		/cultivar="Heinz 1706"
		/db_xref="taxon:4081"
		/clone_id="tox0002617r"
		/clone_id="CUGI Tomato BAC Library"
		/tissue_type="Nuclei preparation from leaf"
		/lab_host="E. coli DH10B"
		/note="Vector: pBelobAC 11; Site.1: HindIII; Site.2: HindIII; Tomato is a vegetable crop that ranks second only to potatoes in value and importance. Among plant geneticists and physiologists, tomato represents an ideal dicot model beside Arabidopsis and monooct rice to derive genomic information from. To facilitate the genome analysis of tomato, we have constructed a tomato BAC library that is suitable for positional cloning, physical mapping, and genome sequencing. The library contains 129 ,000 clones and a random sampling of 498 clones indicated an average insert size of 117.5 kb. With 15x haploid genome equivalents (1C equals 953 Mb) (Armuganathan and Earle, 1991), the probability to recover any particular sequence is greater than 99%. High stability, large insert and ease in manipulation make BAC libraries the choice for genome sequencing. Pre-characterization of a few hundred bases of insert ends will make BAC clones extremely useful for rapid contig assembly (Venter, Smith, and Hood, 1996). Here we present the construction, characterization of the tomato BAC library, and preliminary analysis of the 1536 tomato BAC end sequences."
BASE COUNT	191 a	150 c 96 g 180 t
ORIGIN		
Query Match	31.1%	Score 235.4; DB 13; Length 617;
Best Local Similarity	62.4%	Pred. No. 4.5e-54;
Matches	385; Conservative	0; Mismatches 231; Indels 1; Gaps 1.
84	cctaagaaggaggcattccglttcgltccctaataagaaatgtaatgtccgcgaagg	143
616	cctaagaaaggaggatgcatgctgtgctcccaactaaataaagaactctttccaatgaca	557
144	attatcaactgcatataagatggttgatcgatttcaggaataactaataaagccactaggaa	203
556	cgggttactgagatgagagcggtgtgtatgattaccacaactaaattcagactgaaaaa	497
204	gatactacccttgccttattatgacacatagtctagaagaagttgccaacatcaaacac	263
496	gaccacttttcgatatgccccttcattgacatgacatgttgatgacacttccgaaaaagcg	437
264	tctctgcttctaagacgttatctcaattctcccaataccagttgacaatctgacag	323
436	tacggtttcttgatgagatattccggsgtgaatgacattttctatctcactagaaagca	377
324	gagaaacacacttcaactgccttcggttaccttgccttatagaagctatgacctttg	383
376	tgaataacacacattttactgtccatgtgacacctttgcgttcgaaagaaatgacatttg	317
384	ttatgtaatgacacctccacacttcaaaatgataatgataatcttccctaactttgt	443
316	tttttgcaatgaccccccacacattttcagagatgatatgctcactttctgacatgctg	257
444	gaaacatgtgcgaagtttcatagatgactttccgtaacgaaggtctcttttgatg	503
256	gaagattttcatagaaattttt -atgagatgatttttttctgtgttgatatttcgtgcgg	198
504	tgtctcaagcaacttcatgagctctacagaagttgaagaagccaatctgtcttgaat	563
197	tgtttgacaaattttatctgaggtcttgaagatgtgaaacctgaaattttagactagat	138
564	gaggaagatgccaactttatgttaataagaagatcgtctttagaataaaatttctgaa	623
137	tgggaagaagatcattttcatggtgaaagacggtattgttttgggtatcactattcataa	78

Oy	624		aggagcttcaaacgcataagaagtgtgatgacatcgacaacatgcattaccacaca	683
Dd	77	AAGGAATATAGGTGTGATTCACGTTAAAGTCGAGSTAAATAGACAATTCCGCCACCGATC	18	
Oy	684	galatcaaaagattaag	700	
Dd	17	TTTTATGAAGTGCTGAG	1	
RESULT	11			
LOCUS	BF657025/c			
DEFINITION	BF657025	482 bp	mRNA	EST
ACCESSION	OV2_18.C04.g1_A002	Ovary 2 (OV2)	Sorghum bicolor cDNA,	20-DEC-2000
VERSION	Bf657025		sequence.	
KEYWORDS	Bf657025.1	GI:11922159		
SOURCE	EST.			
ORGANISM	Sorghum.			
	Sorghum bicolor			
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC			
	clade; Panicoideae; Andropogoneae; Sorghum.			
REFERENCE	1 (bases 1 to 482)			
AUTHORS	Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt			
	,L.H.			
TITLE	An EST database from Sorghum: ovaries of varying immature stages			
JOURNAL	Unpublished (2000)			
COMMENT	Contact: Cordonnier-Pratt MM Department of Botany The University of Georgia Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA Tel: 706 542 1860 Fax: 706 542 1805 Email: mmpratt@uga.edu Seq primer: T7 High quality sequence stop: 481 POLYA-No.			
FEATURES				
SOURCE				
	Location=Qualifiers			
	1..482			
	/organism="Sorghum bicolor"			
	/db_xref="taxon:4558"			
	/clone_lib="Ovary 2 (OV2)"			
	/note="Organ: Mix of ovaries of varying immature stages			
	from 8-week-old plants; Vector: pliuscript II from lambda			
	zap II; Site1: XhoI; Site2: EcoRI; The library was made			
	from poly-A RNA in the cloning vector lambda zap II.			
	Clones to be sequenced were prepared by mass excision."			
BASE COUNT	137 a 105 c 87 g 153 t			
ORIGIN				

Query Match	30.1%	Score 227.8	DB 11	Length 482
Best Local Similarity	69.4%	Pred. No. 5.2e-52		
Matches 310	Conservative 0	Mismatches 137	Indels 0	Gaps 0

QY	2	aaaagagctcttgaagctccctgatalgaagatatactatacattgttgcataagcattg	61
Db	448	AAAGCAAGCTTTAAAGATTGCTTCACATGACAGGGATTTATATCTGTGCCCCACAGATGAGTG	389
QY	62	ggtgagtcggtgcataagcgtttcccaagaagaagacattaccgttgcctccaatgata	121
Db	388	GGTGAGCCCCAGTACAAAGTTGTGCCCAAAAAGGAGAGCATGACAGTTATTTATCAATGAAAA	329
QY	122	ggatgaatttgtcccgccaggaagattatatactacgtgctatagatggtgatcgaattcaga	181
Db	328	GAATAGACTTAATTCGGCAACGCACCATACACAGATGGCGGAGTGCATTAAGATACAGAAA	269
QY	182	actgaataaagccactagaagaatcattaccccttgccttttatacgcacatagtctga	241
Db	268	ACTAAATATTAACCGACGAGAAAGATCACTTCCTTACTCTTTATAGATGAATGCTGTA	209
QY	242	aaggttgcacaactcacacaactctgtcttcttcagaagcgttatctttagtttccccaat	301

Db		208	GGCGTTAGCAAAACCATTGCTTCTTTTGTTTTTATGATGATATTCACGGATCATCAAAAT	149
QY		302	accagttgcacaaatctcgatcaggagaagaaacacttcaacttcgccccttggtlactctggc	361
Db		148	CCCGATCCATCCCGATGATCMAAGCAAACACACTTTTACATGCCTTATGTGAACCTTAATGC	89
QY		362	ttaagaacgaagccctcttggtcctaataagcacctgcacaccttccaagaagtatgat	421
Db		88	TATGCTACAATGTCCTTTGGGGTTATGATATGACACACAGCTTCTTTTCMAAGATGATGAT	29
QY		422	ggcatattctctaactttgtgaaaa	448
Db		28	GTCATATTTTCTGTATGATGATGAAGA	2
RESULT_12				
A0966889				
LOCUS			510 bp	DNA
DEFINITION			LERIM76TR LER6 Arabidopsis thaliana genomic clone LERIM76,	SSS
ACCESSION			A0966889	28-JAN-2000
VERSION			A0966889.1	
KEYWORDS			GI:6794590	
SOURCE			SSS.	
ORGANISM			thale cress.	
			Arabidopsis thaliana	
			Eumariophyta; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;	
			Spermatophyta; Magnoliophyta; eudicotyledons: core eudicots:	
			Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsidis.	
REFERENCE			1 (bases 1 to 510)	
AUTHORS			Buell,C.R., Lin,X., Pai,G., Barnstead,M., Bowman,C., Uterbach,T.,	
			Feldbljum,T., Liang,F., Creasy,T. and Fraser,C.M.	
TITLE			Genomic survey sequencing of Landsberg erecta ecotype of	
			Arabidopsis thaliana and identification of sequence-based	
			polymorphisms	
JOURNAL			Unpublished (2000)	
COMMENT			Contact: Xiaoying Lin	
			The Institute for Genomic Research	
			9712 Medical Center Dr., Rockville, MD 20850, USA	
			Tel: 301 838 0200	
			Fax: 301 838 0208	
			Email: atetigr.org	
			For additional information, see http://www.tigr.org/tdb/at/at.html	
			Seq primer: TR	
			Class: shotgun.	

FEATURES		location/Qualifiers
source	1..510	/organism="Arabidopsis thaliana" /strain="Landsberg erecta" /db_xref="taxon:3702" /clone="LERIM76" /clone_lib="LERG"
BASE COUNT	157 a 92 c 96 g 165 t	/note="Organ: Leaf; Vector: pUC19K; Total genomic DNA was sheared to 0.4-0.7 kbp before ligation."
ORIGIN		
Query Match	29.4%; Score 222..2; DB 13; Length 510;	
Best local Similarity	66.3%; Pred. No. 1.8e-50;	
Matches 320; Conservative	0; Mismatches 163; Indels 0; Gaps 0;	
QY	2	aaaagaggttctgaagctcctcgatgaaggtatatactatcaatgttgcctatagcattg 61
Db	27	AAAAGACATACCTTAATATTATAGATGTGGGGATTATCTACCTATCTCGATTAACACTTG 86
QY	62	ggtagagtcggttgatagcgcttcttaagaagggaagcattacggtgtccctaattgataa 121
Db	87	GGTATGCTTACGTTTCATTGGCTTCCTTAATAAAGGAGGTATTAAGAAGTTGTTAAAAACGCTTA 146
QY	122	ggaatgaattgatcccgacgaagatatacaacgcgcgtatagatgtagtgcattcaggaa 181
Db	147	AGATAAACCTATACCCACATAGGACAATTAACATGGAATGATGTATTACTATATAGGAA 206

Oy	182	acgaataaaagccactcagaagaatcatcaaccttgccttttacatcgaccatlatgctaga	241
Db	207	GTTAAATGCTGCCTCTTAGAGAGACCAATTTCCTTCCATTTTCATTAATTCAGATGCTAGA	266
Oy	242	aagattgtccaacaactcaaacacttcgcctttcttagacgglttatctctagtcttcccacaa	301
Db	267	AAGATTAAACAACATGATTCTACTATGGCTCCTAGATGATFATAGCGGATTTTTTCAAA	326
Oy	302	accagttgcacacatcgtatcagaagaagaacacactttcacctgccttcogtactcttg	361
Db	327	CTCATTTTCACCACCAATATGCCAAGAAAACATCACTTTTACTTTCCTTAATGGACCTTTGC	386
Oy	362	ttaagaagatgatgcctttgctttaatgatgcacttgcacacttccaagaatgatgat	421
Db	387	TTCACAGCGTATGCGCTTGCGTCTCGGTATGATGACCTGCTACTCTTCACATGGCGATATAC	446
Oy	422	gactatatctctctaacttttgtgaaaaatattgtcgagglltcaatgatgatcttccgt	481
Db	447	TTCCATTTCCTCGGATTTGATATAGACGATGATGATGATTCATGATGATTTTAAATG	506
Oy	482	tta	484
Db	507	GTA	509
RESULT 13			
LOCUS	A0958534		
DEFINITION	LERRAY15TR LERA Arabidopsis thaliana genomic clone LERRAY15, DNA sequence.		
ACCESSION	A0958534		
VERSION	A0958534.1	GI:6786235	
KEYWORDS	GSS.		
SOURCE	thale cress.		
ORGANISM	Arabidopsis thaliana Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi. 1 (bases 1 to 726)		
REFERENCE	Buell,C.R., Lin,X., Pal,G., Barnstead,M., Bowman,C., Uterbach,T., Feildblum,T., Liang,F., Creasy,T. and Fraser,C.M. Genomic survey sequencing of Landsberg erecta ecotype of Arabidopsis thaliana and identification of sequence-based polymorphisms Unpublished (2000) Contact: Xiaoying Lin The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0200 Fax: 301 838 0208 Email: atel@igr.org For additional information, see http://www.tigr.org/tdb/at.html Seq primer: TR Class: shotgun.		
JOURNAL			
COMMENT			
FEATURES			
source	Location/Qualifiers		
	1..726		
	/organism="Arabidopsis thaliana"		
	/strain="Landsberg erecta"		
	/db_xref="taxon:3702"		
	/clone="LERRAY15"		
	/clone_1fb="LERA"		
	/note="Organ: Leaf; Vector: PHOSI; Total genomic DNA was sheared to 0.9-1 Kbp before ligation."		
BASE COUNT	230 a 142 c 133 g 221 t		
ORIGIN			
Query Match	29.2%; Score 221.2; DB 13; Length 726;		
Best Local Similarity	71.2%; Pred. No. 3.7e-50;		
Matches	292; Conservative 0; Mismatches 118; Indels 0; Gaps		
Oy	2	aaaagaagtgctgaagctccctgataaggtatcatactaatcagtgtgctcatagcgattg	61

Db	Accession	Version	Keywords	Organism	Source	Location/Qualifiers
Db	317	AAAAAATTTTGA	ACTGCTGATGCTGGTGATCATACCAATCTCATAGTACTTG	376		
Qy	62	ggtagtcggtgatagacggttccttaagaaggagacattacggtgtccctaatgataa	121			
Db	377	GGTTTTCCAGTGCATGCGTCCTAA	AAAAAGGCGAATGACGTGTGTCAAAATGAAAA	436		
Qy	122	ggatgaattgattcccgacagagatattcaacgctcattagagtgatgtatccaagaa	181			
Db	437	AGATGAACCTGATCCCTACTAGAACTATTAACCTGGTCATAGAAATGTCATAGATTATAGAA	496			
Qy	182	actgaataaagccacactagaagaagatacattaccccttgcccttllatcgacatatgata	241			
Db	497	GTTGATCTCTGCATCTAGGAAGAATTCATTTCTTTACCATTCATAGACCAAAATGCTTGA	556			
Qy	242	aaggtgtccaaactcacacacactctcgtcttcttagaaggttattctatggttccccaat	301			
Db	557	ACGATTACCTATCATCATCATATTTATTTGCTTTCTTGATGGATACAGTGGTTCTTTCAAAAT	616			
Qy	302	accaggttcacacatcgtatcgtagggagaaacacacttaccgctccgcttcgttaccattc	361			
Db	617	ACCAATTCACCCCTAATGATGACAGAAAAACCACTTTCACGTGTCCTTATGGAACCTTTTGC	676			
Qy	362	ttatagacgtatgccttcttgctattgatagcacccctgcaccccttccaa	411			
Db	677	CTATAGAGANTGGCATTTGGTTTATGCAATGCTGCTGCAACATTTTGA	726			
RESULT	14					
LOCUS	AO967852	510 bp	DNA	GSS	28-JAN-2000	
DEFINITION	LERIV29TRB	LERG	Arabidopsis thaliana genomic clone LERIV29, DNA			
ACCESSION	AO967852					
VERSION	AO967852.1	GI:6795553				
KEYWORDS	GSS.					
ORGANISM	thale cress.					
	Arabidopsis thaliana					
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;					
	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;					
	Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.					
REFERENCE	1 (bases 1 to 510)					
AUTHORS	Buell,C.R., Lin,X., Pal,G., Barnstead,M., Bowman,C., Uterbach,T.,					
	Feldblyum,T., Liang,F., Greasy,T. and Fraser,C.M.					
	Genomic survey sequencing of Landsberg erecta ecotype of					
	Arabidopsis thaliana and identification of sequence-based					
	polymorphisms					
JOURNAL	Unpublished (2000)					
COMMENT	Contact: Xiaoying Lin					
	9712 Medical Center Dr., Rockville, MD 20850, USA					
	Tel: 301 838 0200					
	Fax: 301 838 0208					
	Email: ateligr.org					
	For additional information, see http://www.tigr.org/tdb/at.html					
	Seq primer: TR					
	Class: Shotgun.					
FEATURES	Location/Qualifiers					
	1..510					
	/organism="Arabidopsis thaliana"					
	/strain="Landsberg erecta"					
	/db_xref="taxon:3702"					
	/clone="LERIV29"					
	/clone_lib="LERG"					
	/note="Organ: Leaf; Vector: pUC19/K; Total genomic DNA was					
	sheared to 0.4-0.7 kbp before ligation."					
BASE COUNT	172 a 104 c 91 g 142 t					
ORIGIN						
Query Match	28.4%;	Score 214.8;	DB 13;	Length 510;		
Best local Similarity	69.5%;	Pred. No. 2e-48;				
Matches 291; Conservative	0;	Mismatches 128;	Indels 0;	Gaps 0;		

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 1, 2002, 21:29:33 ; Search time 79.08 Seconds
(Without alignments)
235.109 Million cell updates/sec

Title: US-09-586-106-63
Perfect score: 1347
Sequence: 1 KEVVKLDGEGIIYVHAHSDM.....INRFLGSGGFEYRRIKDFTK 251

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: /SID2/gcgdata/geneeq/geneeq/AA1980.DAT:*
2: /SID2/gcgdata/geneeq/geneeq/AA1981.DAT:*
3: /SID2/gcgdata/geneeq/geneeq/AA1982.DAT:*
4: /SID2/gcgdata/geneeq/geneeq/AA1983.DAT:*
5: /SID2/gcgdata/geneeq/geneeq/AA1984.DAT:*
6: /SID2/gcgdata/geneeq/geneeq/AA1985.DAT:*
7: /SID2/gcgdata/geneeq/geneeq/AA1986.DAT:*
8: /SID2/gcgdata/geneeq/geneeq/AA1987.DAT:*
9: /SID2/gcgdata/geneeq/geneeq/AA1988.DAT:*
10: /SID2/gcgdata/geneeq/geneeq/AA1989.DAT:*
11: /SID2/gcgdata/geneeq/geneeq/AA1990.DAT:*
12: /SID2/gcgdata/geneeq/geneeq/AA1991.DAT:*
13: /SID2/gcgdata/geneeq/geneeq/AA1992.DAT:*
14: /SID2/gcgdata/geneeq/geneeq/AA1993.DAT:*
15: /SID2/gcgdata/geneeq/geneeq/AA1994.DAT:*
16: /SID2/gcgdata/geneeq/geneeq/AA1995.DAT:*
17: /SID2/gcgdata/geneeq/geneeq/AA1996.DAT:*
18: /SID2/gcgdata/geneeq/geneeq/AA1997.DAT:*
19: /SID2/gcgdata/geneeq/geneeq/AA1998.DAT:*
20: /SID2/gcgdata/geneeq/geneeq/AA1999.DAT:*
21: /SID2/gcgdata/geneeq/geneeq/AA2000.DAT:*
22: /SID2/gcgdata/geneeq/geneeq/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	999	74.2	1801	21	AAV32434
2	790	58.6	200	21	AAV32431
3	425	31.6	1217	20	AAV14565
4	225.5	16.7	916	21	AAE03130
5	219.5	16.3	1871	21	AAE03133
6	219.5	16.3	1871	21	AAE03133
7	217.5	16.1	1871	20	AAV57165
8	197.5	14.7	1156	21	AAV01078
9	195.5	14.5	401	18	AAV01078
10	195.5	14.5	401	18	AAV01078
11	195.5	14.5	401	18	AAV01078
12	190.5	14.1	1194	18	AAW39272
13	189.5	14.1	1145	18	AAW32097
14	189.5	14.1	1145	18	AAW32097
15	174.5	13.0	1784	11	AAW05898
16	174.5	13.0	1784	11	AAW05898
17	169	12.5	740	12	AAE1516
18	167.5	12.4	1203	14	AAW81572
19	165.5	12.3	524	20	AAE1516
20	164.5	12.2	524	21	AAE1516
21	163.5	12.1	1196	16	AAW75189
22	162.5	12.1	672	21	AAE12457
23	158.5	11.8	429	19	AAW68403
24	158.5	11.8	429	19	AAW68403
25	158.5	11.8	498	21	AAE03349
26	158.5	11.8	504	16	AAW27322
27	158.5	11.8	504	18	AAW27322
28	158.5	11.8	504	21	AAE03350
29	158.5	11.8	1079	20	AAE03163
30	158.5	11.8	1199	21	AAE12994
31	158.5	11.8	1224	20	AAE17947
32	158.5	11.8	1737	21	AAE10044
33	151.5	11.2	1197	22	AAE49468
34	134	9.9	1053	11	AAE05614
35	130	9.7	768	19	AAW71062
36	126	9.4	372	22	AAE81066
37	124.5	9.2	1014	9	AAE80810
38	124.5	9.2	1027	13	AAE1773
39	124.5	9.2	1036	13	AAE20599
40	124.5	9.2	1036	21	AAE1978
41	123.5	9.2	1055	18	AAE13055
42	123.5	9.2	3080	10	AAE93285
43	123.5	9.2	3210	9	AAE1771
44	123	9.1	796	18	AAW36024
45	123	9.1	796	19	AAW71080

ALIGNMENTS

RESULT 1	
AAV32434	
ID	AAV32434 standard; Protein: 1801 AA.
XX	
AC	AAV32434;
XX	
DT	27-MAR-2000 (first entry)
XX	
DE	Plant generic retroelement.
XX	
KW	Retroelement; retrovirus; transgenic plant; gene transfer;
KW	soybean; pea; calypso; athila; cyclops.
XX	
OS	Glycine max.
OS	Arabidopsis thaliana.
OS	Pisum sativum.
XX	
PN	WO9960842-A2.
XX	
PD	02-DEC-1999.
XX	
PF	28-MAY-1999; 99WO-US11858.
XX	
PR	29-MAY-1998; 98US-0087125.
XX	
PR	28-MAY-1999; 99US-0087125.
XX	
PA	(WRIG/) WRIGHT D A.
XX	
PA	(VOYT/) VOYTAS D F.
XX	
PI	Wright DA, Voytas DF.
XX	
DR	WPI: 2000-105586/09.
XX	
DR	N-PSDB; AA234261.
XX	

Porcine retrovirus
Miniature swine re
Retroviral protein
Gene product of fl
Pely F6a provirus
Soybean chlorotic
Nus dunni endogeno
Reverse transcript
Osteoinductive ret
Moloney murine leu
Fely-A pol protein
Novel reverse tran
M-MLV reverse tran
Novel reverse tran
Reverse transcript
M-MLV reverse tran
MLV reverse trans
MOMLV pol gene pro
MMLV gag-pol prote
Canine retrovirus
BIV pol gene produ
Multiple sclerosis
HIV protease and r
Sequence of pol pr
ROD HIV-2 polymera
HIV-2 ROD isolate
HIV-2 provirus-enc
Sequence of clone
Deduced sequence e
MSRV-1 pol protein
Multiple sclerosis

10
11
12
13
14
15

KW Genome; retroelement; retrovirus; ZAM; gag; pol; env; LTR; vector;
KM long terminal repeat; gene therapy.
KW
XX

OS Drosophila melanogaster.
XX
PN FR2772045-A1.
XX
PD 11-JUN-1999.
XX
PF 10-DEC-1997; 97FR-0015655.
XX
PR 10-DEC-1997; 97FR-0015655.
XX
PA (UYAU-) UNIV AUVERGNE.
XX
PI Leblanc P, Vauray C;
XX
DR WPI; 1999-359998/31.
DR N-PSDB; AAX83445.
XX
PT New Drosophila ZAM retroelement nucleic acid - useful in gene
PT transfer or gene therapy
XX
PS Disclosure; Fig 1; 55pp; French.
XX
CC This sequence represents the pol protein from a novel retroelement ZAM
CC found in Drosophila melanogaster. Similarly to other retroelements
CC and retroviruses, ZAM contains 3 genes (gag, pol and env) and long
CC terminal repeats (LTR's) at either end of the genome. Vectors containing
CC the ZAM nucleic acid sequence can be used for gene transfer or gene
CC therapy.
XX
SQ Sequence 1217 AA;

Query Match	31.6%	Score 425	DB 20	Length 1217
Best Local Similarity	37.5%	Pred. No. 8.9e-37		
Matches 94; Conservative	46	Mismatches 97	Indels 14	Gaps 3

[illegible]

RESULT	4
AAB03130	
ID	AAB03130 standard; Protein; 916 AA

DT	10-OCT_2000 (first entry)
XX	
DE	Gag-pol readthrough protein encoded by <i>C. albicans</i> retrotransposon 15.
XX	
KW	retrotransposon; pCal; TCA2; TY1; copia; long terminal repeat; LTR
KW	gag gene; group antigen; polyploidin; pol; aspartate protease; integrase
KW	reverse transcriptase; RNaseH; pseudoknot; readthrough translation;
KW	stop codon suppression; gene delivery; gene therapy vector;
KW	genetic vaccine composition; immunogenic; transgenic animal.

OS	Candida albicans.
xx	
PN	WO2000026397-A1.
xx	
PD	11-MAY-2000.
xx	
PF	01-NOV-1999; 99WO-NZ00179.
xx	
PR	30-OCT-1998; 98CA-2249046.
PR	30-OCT-1998; 98US-0106342.
xx	
PA	(JANC) JANSSEN PHARM NV.
xx	
PI	Luyten WHML, De Backer MD, Neijssen BJM, Poulter RTM;
xx	
DR	WPI; 2000-365640/31.
DR	N-PSDB; AAA57953.
xx	
PT	Novel retrotransposon expression vectors useful for expressing an
xx	antigen, epitope or therapeutic agent, or detecting genes or the
PT	presence of Candida in a sample -

Example 12; Fig 33; 204pp; English.

The invention relates to novel retrotransposons from the yeast *Candida albicans* which have a copy number of 40-150, preferably 50-100 copies per genome. In particular, the invention relates to the novel *C. albicans* TY1/copia retrotransposon *pcal* (AA557920), and to the integrated form of this retrotransposon, designated Tca2, and to the novel *C. albicans* retrotransposons 1-28. *pcal* was initially isolated from *C. albicans* HOG1042 and has a copy number of 50-100 copies per cell. It comprises identical 280 bp long terminal repeats (LTRs) and two open reading frames (ORFs). The first ORF encodes a gag (group antigen) protein, and the second ORF encodes a polyprotein (pol) consisting of an aspartate protease, integrase, reverse transcriptase (RT) and RNaseH. The gag and pol ORFs of *pcal* are in the same reading frame, separated only by a termination codon (TGA). Translation of the pol ORF occurs through the occasional readthrough suppression of a pseudoknot within the gag-pol mRNA. The retrotransposons of the invention can be used as vectors for *in vitro* or *in vivo* transformation and expression. They can thus be used for the delivery and expression of a therapeutic, immunological or immunogenic molecule (e.g., an antigen) and may also be used for eliciting an immunological response in a host organism. They are therefore useful in genetic vaccine compositions and for gene therapy, particularly where the use of retroviral vectors is unsafe or undesirable. Additionally, the retrotransposons may be used to generate transgenic animals, to detect the presence of *Candida* in a sample, to detect and disrupt genes, and to assign functions to nucleotide sequences. Sequences AA03129-B0313 represent proteins or protein fragments encoded by the novel *C. albicans* retrotransposons of the invention.

Query Match	16.7%;	Score 225.5;	DB 21;	Length 916;
Best Local Similarity	24.2%;	Pred. No. 2.3e-15;		
Matches 61;	Conservative	112;	Indels 31;	Gaps 6

```

QY      7 LDGGIIYVAHSPWSPVSHVSKGGIITVPDPDKDELLIPORIIITGYRRVDFDKLNAT- 65
      11 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     273 leagvlyppadiawlhlplfpirktna-----nq-----sstklaovdlrrlnkvtv 317
      11 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY     66 RKCHYPLPFIIDHMLERLSKTLHFCELDGYSSFSQIPVAOSDOEKRTTTPCEPGTAYRRMP 125
      11 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db    318 rmltvpdctkd-llslstcdshfalsadlknaatyshkdsikyfgjstsegunyctflpl 376
      11 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY    126 FGLCNAPAFQRCMAIIFSNFCENIVE-----VMDDFSVYGGSSFDDCLSNDRVQRC 179
      11 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db    377 fgainsp-----tlfnfvrqtliegjpcflfymdallthkthlmdhmlrrimekl 428
      11 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY    180 KDNLVLNGKCHFMVNEGIVLGHKTSSEGVIEVDKAKYDALDKMYPPIPDIGINSPLGHG 239

```

```

Db      429 nehqfymynkmqllttkinflysgankispdiskigalqmwelptttqirafvfnfs 488
QY      240 GFYRRFKDFPK 251
XX      489 nhfriflpeiaak 500

RESULT 5
AAB03133
ID      AAB03133 standard; Protein; 916 AA.
XX
AC      AAB03133;
XX
DT      10-OCT-2000 (first entry)
DE
DE      Pol fragment 1 encoded by 4545 bp Candida albicans retrotransposon 14.
XX
KM      Retrotransposon; pcal; Tca2; Tyl; copia; long terminal repeat; LTR;
KM      gag gene; group antigen; polyprotein; pol; aspartate protease; integrase;
KM      reverse transcriptase; Rnaseh; pseudoknot; readthrough translation;
KM      stop codon suppression; gene delivery; gene therapy vector;
KM      genetic vaccine composition; immunogenic; transgenic animal.
XX
OS      Candida albicans.
XX
PN      WO200026397-A1.
XX
PD      11-MAY-2000.
XX
PF      01-NOV-1999; 99MO-NZ00179.
XX
PR      30-OCT-1998; 98CA-2249046.
PR      30-OCT-1998; 98US-0106342.
XX
PA      (JANC) JANSSEN PHARM NV.
XX
PI      Luyten WHML, De Backer MD, Nelissen BJM, Poultier RTM.
DR      WPI: 2000-365640/31.
DR      N-PSDB: AAA57995.
XX
PT      Novel retrotransposon expression vectors useful for expressing an
PT      antigen, epitope or therapeutic agent, or detecting genes or the
PT      presence of Candida in a sample.
XX
PS      Disclosure: Fig 71; 204pp; English.
XX
CC      The invention relates to novel retrotransposons from the yeast Candida
CC      albicans which have a copy number of 40-150, preferably 50-100 copies
CC      per genome. In particular, the invention relates to the novel C.
CC      albicans Tyl1/copia retrotransposon pcal (AA57920), and to the integrated
CC      form of this retrotransposon, designated Tca2, and to the novel C.
CC      albicans retrotransposons 1-28. pcal was initially isolated from C.
CC      albicans HOG1042 and has a copy number of 50-100 copies per cell. It
CC      comprises identical 280 bp long terminal repeats (LTRs) and two open
CC      reading frames (ORFs). The first ORF encodes a gag (group antigen)
CC      protein, and the second ORF encodes a polyprotein (pol) consisting of an
CC      aspartate protease, integrase, reverse transcriptase (RT) and Rnaseh. The
CC      gag and pol ORFs of pcal are in the same reading frame, separated only by
CC      a termination codon (TGA). Translation of the pol ORF occurs through
CC      the occasional readthrough suppression of the stop codon, which is
CC      mediated by the formation of a pseudoknot within the gag-pol mRNA.
CC      The retrotransposons of the invention can be used as vectors for in
CC      vitro or in vivo transformation and expression. They can thus be used
CC      for the delivery and expression of a therapeutic, immunological or
CC      immunogenic molecule (e.g., an antigen) and may also be used for
CC      eliciting an immunological response in a host organism. They are
CC      therefore useful in genetic vaccine compositions and for gene therapy,
CC      particularly where the use of retroviral vectors is unsafe or
CC      undesirable. Additionally, the retrotransposons may be used to generate
CC      transgenic animals, to detect the presence of Candida in a sample, to
CC      detect and disrupt genes, and to assign functions to nucleotide
  
```

```

CC      sequences. Sequences AAB03129-B03139 represent proteins or protein
CC      fragments encoded by the novel C. albicans retrotransposons of the
CC      invention.
XX
SQ      Sequence 916 AA:

Query Match 16.7%; Score 225.5; DB 21; Length 916;
Best Local Similarity 24.2%; Pred. No. 2.3e-15;
Matches 61; Conservative 48; Mismatches 112; Indels 31; Gaps 6;

QY      7 LDEGIIVHANSDWSPVHSVPKKGITVVPNDKDELIPQRTITGIRAVIDPRKINKAT- 65
DB      273 leagylvpapdawlhpfirktua-----nq-----sstklavdlrrinkvlt 317
QY      66 RKDHPLPFDHMLERLSKLTNFCFLDGYSSFSQIPVAOSDQETTTCTPCGTAYARRMP 125
DB      318 rmytptldtkd-lissldshyfsaldlknafgyvskhkslkfygistsgnyfttlp 376
QY      126 FGLCAPATPQRCMAAIPSNFCENIVE-----VEMDFSVYGSFDDCLSNLRVLDRC 179
DB      377 fgalsp-----tlfntrqqllegipclflymdllthkclthdmslrrimeki 428
QY      180 KDTNVLNGEKCHPMVNEGIVLGHKISERGETVDKAKVDAIDKMPYPTDIDKINSFLGHG 239
DB      429 nehqfymynkmqllttkinflysgankispdiskigalqmwelptttqirafvfnfs 488
QY      240 GFYRRFKDFPK 251
DB      489 nhfriflpeiaak 500

RESULT 6
AAB15671
ID      AAB15671 standard; Protein; 1871 AA.
XX
AC      AAB15671;
XX
DT      01-DEC-2000 (first entry)
DE
DE      Sugarcane bacilliform virus amino acid sequence.
XX
KM      Sugarcane bacilliform virus; SCBV; SCBV promoter; transgenic plant.
XX
OS      Sugarcane bacilliform virus.
XX
PN      US6093569-A.
XX
PD      25-JUL-2000.
XX
PF      08-JUL-1999; 99US-0349546.
XX
PR      09-AUG-1996; 96US-0694869.
XX
PA      (MIND) UNIV MINNESOTA.
XX
PI      Lockhart B, Torbert K, Olszewski N, Tzafir I, Somers DA.
DR      WPI: 2000-514111/46.
DR      N-PSDB: AAA73991.
XX
PT      Producing transformed plant cells useful for creating transgenic plants
PT      for commercial, research and agricultural purposes, comprises
PT      introducing a recombinant DNA comprising a sugarcane bacilliform
PT      promoter.
XX
PS      Disclosure: Fig 1; 34pp; English.
XX
CC      The present sequence is the deduced amino acid sequence of sugarcane
CC      bacilliform virus (SCBV). The promoter of the gene encoding this protein
CC      may be used to express proteins and RNA transcripts in transgenic plants.
CC      The promoter is operably linked to the desired nucleotide sequence and
CC      the recombinant DNA is introduced into plant cells. The method is useful
  
```

CC for producing plants having unique signature sequences or other marker
CC sequences which can be used to identify proprietary lines or varieties,
CC and transgenic plants for commercial and research purposes. Transgenic
CC plants may be created for use in traditional agriculture and for use in
CC the manufacture of proteins or other compounds, which are extracted or
CC purified from plant parts or seeds. They may also be used in commercial
CC breeding programmes, including in the creation of new mutant plants
CC through insertional mutagenesis.

Sequence 1871 AA;

Query Match	16.3%;	Score 219.5;	DB 21;	Length 1871;
Best Local Similarity	30.2%;	Pred. No. 2.8e-14;		
Matches	61;	Conservative	37;	Mismatches 99;
			Indels	5;
			Gaps	3;

[illegible]

RESULT 7
AAV57165
ID AAV57165 standard; Protein; 1871 AA.

AC	AAY57165;
XX	
DT	11-FEB-2000 (first entry)

DE Amino acid sequence of sugarcane bacilliform virus gene.

KW Sugarcane bacilliform virus; ScBV; promoter; transgenic plant;
monocot plant; dicot plant.
KW

OS Sugarcane bacilliform virus.

PN US5994123-A.

PD 30-NOV-1999.

PF 09-AUG-1996; 96US-0694869.

PR 09-AUG-1996; 96US-0694869.

PA (MINN) UNIV MINNESOTA.

PI lockhart B, Torbert K, Olszewski N, Tzatlir I, Somers DA,

DR WPI; 2000-038263/03

XX
XX
T 01 30 00 DNA COER

PT proteins and RNA transcripts in transgenic plants -

PS Disclosure; Fig 1A-B; 34pp; English.

CC The invention provides sugarcane bacilliform virus (SCBV) promoters
CC selected from sequences shown in AA239511-513. The promoter is used to
CC express proteins and RNA transcripts in transgenic plants in order to
CC improve their characteristics. The promoter confers constitutively high
CC levels of expression of operably linked DNA sequences in

CC both monocot and dicot plants, plant tissue, plant parts or plant cells.
CC The present sequence represents the sugarcane bacilliform virus amino
CC acid sequence.

Sequence 1871 AA;
SQ

Query Match	16.3%;	Score 219.5;	DB 21;	Length 1877;
Best Local Similarity	30.2%;	Pred. No. 2.8e-14;		
Matches	61;	Conservative	37;	Mismatches 99;
			Indels	5;
			Gaps	3

QY	53	RMVDEPRKLNKATKPKKHDPYLPPEFDNHLERISKTHTHCEFDGYSSTFSQJIPVAKSODQEKTFE	112
Db	1316	rlvlnvyrklndctkwpdysltpgjnallkvnarekisktsfdlksqfhyvameeelsipltaf	1375
QY	113	TCPGFTFAYRRMRPGLCNABATGFORCMMATFSNFCGENIVEFMDESVYSGSFDDCSNLT	172
Db	1376	saynelYewlvmprgjlknapaqlfqrmqdcfgfr-tcgflavylld11vtsedeeghaeh1	1434
QY	173	DRVLQRCRDNTMLVNGEKCFHVNNEGVYGHKISESGIVSD--AKVDAIDKMPYPYDI	229
Db	1435	wkmlqickrtnglllspskyrklygvkvdfvglsctlgdqlavgenllkkaiefdekltk-	1493
QY	230	KGIRSLFGHGGEVRRFIDKFTK	251
Db	1494	eglkswlatltnyarnhikdngk	1515

RESULT	8
AA01078	
ID	AA01078 standard; Protein; 1871 AA

AC AAY01078;

DT 08-JUN-1999 (first entry)

DE Sugarcane bacilliform virus protein sequence

KW Sugarcane bacilliform virus; promoter; SCBV promoter; transgenic plant;
KW water deficit resistance; pest resistance; herbicide resistance;
KW nutritional content; protein manufacture.

OS Sugarcane bacilliform virus.

PN WO9909190-A1.

PD 25-FEB-1999.

PF 13-AUG-1997; 97WO-IB01338

PR 13-AUG-1997; 97WO-IB01338

PA (MINN) UNIV MINNESOTA.

PI Lockhart B, Olszewski N, Somers DA, Torbert K, Tzafir L,

DR WPI; 1999-181047/15.

2046
XX
XX

PT Newly isolated DNA molecule comprising a sugarcane bacilliform virus
PT promoter - useful for generating fertile transgenic plants with
PT beneficial agronomic characteristics

PS Disclosure; Fig 1; 67pp; English.

CC This sequence represents a sugarcane bacilliform virus protein.
CC
CC The invention relates to a preselected DNA segment, consisting of
CC a sugarcane bacilliform virus (ScBV) promoter (including a biologically
CC active subunit). Fertile transgenic plants produced by transformation
CC with the DNA sequence are useful for conventional plant breeding
CC programs. Other transgenic plants are useful for generating beneficial
CC agronomic characteristics, including resistance to water deficit, pest
CC resistance, herbicide resistance, increased yield, improved nutritional

CC content, and improved processing characteristics. They are also useful
 CC for commercial protein manufacture. The scav promoter is a strong,
 CC constitutive promoter in many cells, enabling the expression of genes in
 CC fertile transgenic monocot or dicot plants, which allows genetic
 CC engineering of improved plant characteristics.

XX Sequence 1871 AA:

Query Match 16.1%; Score 217.5; DB 20; Length 1871;
 Best Local Similarity 29.7%; Pred. No. 4.7e-14;
 Matches 60; Conservative 38; Mismatches 99; Indels 5; Gaps 3;

QY 53 RMVIDFRKLNKATRKRDHYPLFDHMLERLSKLTTHFCFLDGYSSFSQIPVAOSDQEKTF 112
 Db 1316 rlvfnykrindntwppdgsllpglnallknvarakifskfdlksghyqamdeesipltaf 1375
 QY 113 TPCPTGAVRRMPFGCLCNAPATFORCMAAIFSNFCENIVEVPMDFSVYSSFDCLSNL 172
 Db 1376 saynellyewlmpiglknapalifqrkmdqclrg- tegfiavylldllvfseeeqhaehl 1434
 QY 173 DRVLQRCRDTNLVNGEKFHMFVNEGIVLGHKISERGIEVDK---AKYDAIDKMPYPTDI 229
 Db 1435 wkmqlqickrkgllispskylgkvkvdflgstiganglavqenllkkaefdekliktk- 1493
 QY 230 KGRSFLGSGGCFRRRTKDFTK 251
 Db 1494 egikswlatlnyarnhikdmqk 1515

RESULT 9

AA10104
 ID AAB10104 standard; Protein; 1156 AA.

AC AAB10104:
 DT 10-NOV-2000 (first entry)
 XX
 XX

DE Feline foamy virus pol protein.

KW Retroviral vector; anti-FIV; anti-HIV; vaccine; vaccination; cat;
 KW feline immunodeficiency virus; human immunodeficiency virus; FIV; HIV;
 KW pol protein.

OS Feline foamy virus.

PN DE19658441-A1.

PD 21-JUN-2000.

PF 17-DEC-1998; 98DE-1058441.

PR 17-DEC-1998; 98DE-1058441.

PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.

PI Fluegel R, Loechelt M, Flower R, Winkler I;

DR WPI: 2000-432420/38.

DR N-PSDB: AAA40381.

PT Feline foamyvirus retroviral vector for expression of foreign DNA in
 PT animals, especially for vaccination of cats and humans against feline
 PT or human immunodeficiency virus (FIV) or (HIV), respectively -

PS Example 1; Fig 2; 11pp; German.

CC This invention describes a novel retroviral vector to introduce a
 CC desired, expressible DNA into an animal cell, where the retroviral
 CC vector comprises a first DNA sequence, which corresponds to a part of
 CC the reverse transcript of a feline foamy virus (FeV), and a second DNA
 CC sequence permitting propagation in bacteria. The products of the
 CC invention have anti-FIV and anti-HIV activity and can be used in

CC vaccines. The retroviral vector can be used to produce a medicament for
 CC vaccination, in particular for vaccination of cats against feline
 CC immunodeficiency virus (FIV) or humans against human immunodeficiency
 CC virus (HIV). This sequence represents the feline foamy virus pol protein
 CC which is described in the method of the invention.

XX Sequence 1156 AA:

Query Match 14.7%; Score 197.5; DB 21; Length 1156;
 Best Local Similarity 25.9%; Pred. No. 3.5e-12;
 Matches 65; Conservative 47; Mismatches 108; Indels 31; Gaps 6;

QY 6 LLDDEITIHVAHSDMVSVSHVSKKGGITVVPNDKDELIPQRIINGYRWVIDFRKLNKAT 65
 Db 182 llkqgvllq-kestlntrvypvk-----pngr-----wrmvldyrvavkvvt 222
 QY 66 RKDHYPLFPIDH-----MLERLSKLTTHFCFLDGYSSFSQIPVAOSDQEKTFCTCFGFA 120
 Db 223 -----pllavqngshyglilgslfkgryktldlslngfwhplvpdywltafwqkqyc 277
 QY 121 YRRMPFGCLCNAPATFORCMAAIFSNFCENIVEVPMDFSVYSSFDCLSNLDRVLQRC 180
 Db 278 wtvlpqgflnsqglftgvdvdlggipa--veyvdyvishsekenleyldlfnrlk 335
 QY 181 DTNLVNGEKFHMFVNEGIVLGHKISERGIEVDKAVDAIDKMPYPTDIKIRSFHG 240
 Db 336 eaaylislkksnlanslvdfilgltnegrgltdltfxeklenltaptlkqslglln 395
 QY 241 FYRRFIRKDFTK 251
 Db 396 farnfipdfte 406

RESULT 10

AAW32092
 ID AAW32092 standard; Protein; 401 AA.

AC AAW32092:

DT 09-FEB-1998 (first entry)

DE Porcine retrovirus partial POL protein from defective virus.

KW Retrovirus; porcine; POL protein; xenotransplantation; infectious;
 KW provirus; organ transplant; donor; activated virus; PCR.

OS Porcine retrovirus.

FN Key Location/Qualifiers

FT Protein 1..401

FT /label= POL.protein

FT /note="Fragment containing putative partial POL viral

FT protein"

FN WO9721836-A1.

PD 19-JUN-1997.

PF 13-DEC-1996; 96WO-US19680.

PR 14-DEC-1995; 95US-0572645.

PA (GEHO) GEN HOSPITAL CORP.

PI Fishman JA;

DR WPI: 1997-332804/30.

DR N-PSDB: AAT74883.

PT New nucleic acid from porcine retroviruses - used for detecting
 PT viruses in transplant or other tissue and for assessing risk of
 PT transmitting infection to graft recipient

XX Claim 16; Fig 2; 128bp; English.

PS This sequence represents a putative partial viral POL protein from a
 CC defective purified swine retrovirus found in PK-15 cells. There are a
 CC few in frame stop codons and apparent frame shifts in the given CDNA
 CC coding sequence which alter features of the translation
 CC (see AAM32093-W32095 for other parts of the sequence). This sequence and
 CC PCR fragments generated from the sequence (see AAT74812-T74882) can be
 CC used to screen organs for porcine retroviruses prior to
 CC xenotransplantation. Transplantation can increase the likelihood of
 CC retroviral activation if intact and infectious proviruses are present.
 CC The porcine retroviral sequence can be used to generate probes to
 CC determine the level (e.g. copy number) of intact (i.e. potentially
 CC replicating) porcine provirus sequences in a strain of xenograft
 CC transplantation donors. It can be used to detect mutations, genetic
 CC lesions or viral recombinants and also to determine the histological
 CC localization of activated retrovirus. Using Polymerase Chain Reaction DNA
 CC Quantitation (PQ) on blood mononuclear cells, infectivity titration and
 CC susceptibility testing can be performed. Ultimately animal donors without
 CC intact porcine retroviral sequences or a lower copy number of viral
 CC elements could be selected.

XX SQ Sequence 401 AA;

Query Match 14.5%; Score 195.5; DB 18; Length 401;
 Best Local Similarity 26.0%; Pred. No. 1.3e-12;
 Matches 67; Conservative 42; Mismatches 120; Indels 29; Gaps 6;

QY 3 VVKLLDEGIIYVAHSDWSPVHSVPKKGITVVPNDKDELIPQRIITGYRWIDPRKLN 62
 DB 147 VQRLIQGIIIVPV-qspwnrlilpvtkpg-----tnd-----yrpqdlrevn 188
 QY 63 KATRKMDYRLPFDHMLERL-SKLTGFCFLDGYSSFSQIPVAOSDDEKTF-----TCP 115
 DB 189 kvqgdihprtpnpyllsalprerwnytlvldkdaefclrlhpsqplffewrptgtgr 248
 QY 116 FGTFAVRMPFGICNAPATQRCMAIFSNFCENIVEV---FMDDPSYVGSFDDCLSN 171
 DB 249 tggltwrlrpgfknspflfdeahlrldanfrlqhpvtllgyvddlllagatkqdcleg 308
 QY 172 LDRVLOKCKDTNLVNEKCHFMWNEGIVLGHKISEKGEVDAKAKYDAIDKMPYPTDIK 231
 DB 309 tkaillelsdlygrasakkaqicrrevlylgyrlgqgtwltearkktvwqipaptakg 368
 QY 232 IRSFLGHGFGFYRRFIKDF 249
 DB 369 vreflgtagfcrllwipgf 386

RESULT 11

AAB73283
 ID AAB73283 standard; Protein; 401 AA.

XX AAB73283;

XX 23-MAY-2001 (first entry)

XX Defective retroviral genome protein #2 isolated from PK-15 cell line.

XX Retrovirus; graft transplantation; xenotransplantation; PK-15 cell line.

XX Unidentified.

XX US6190861-B1.

XX 20-FEB-2001.

XX 13-DEC-1996; 96US-0766528.

XX 14-DEC-1995; 95US-0572645.

XX

PA (GEHO) GEN HOSPITAL CORP.

XX Fishman JA;

XX WPI; 2001-256211/26.

XX N-PSDB; AAF77726.

XX Assessing risk of endogenous retroviruses in clinical practice and in
 PT xenotransplantation, comprises using probe sequences derived from swine
 PT or miniature swine retroviral genome

XX Disclosure; Fig 2; 127bp; English.

XX The present invention relates to a method for screening a cell or tissue
 CC for the presence or expression of a retrovirus (RV), comprising
 CC contacting a target nucleic acid from the cell or tissue with a second
 CC nucleic acid from the present invention (e.g. AAF77726 or a fragment
 CC thereof). The method is useful for RV detection and to assess graft
 CC transplantation risk. Screening of animals allows the elimination of
 CC donors with active replication of known viruses. Inactive proviruses can
 CC be detected and inactivated, allowing identification and elimination of
 CC potential human pathogens derived from swine in a manner not possible in
 CC the outbred human organ donor population and is important to the
 CC development of human xenotransplantation.

XX SQ Sequence 401 AA;

Query Match 14.5%; Score 195.5; DB 22; Length 401;
 Best Local Similarity 26.0%; Pred. No. 1.3e-12;
 Matches 67; Conservative 42; Mismatches 120; Indels 29; Gaps 6;

QY 3 VVKLLDEGIIYVAHSDWSPVHSVPKKGITVVPNDKDELIPQRIITGYRWIDPRKLN 62
 DB 147 VQRLIQGIIIVPV-qspwnrlilpvtkpg-----tnd-----yrpqdlrevn 188
 QY 63 KATRKMDYRLPFDHMLERL-SKLTGFCFLDGYSSFSQIPVAOSDDEKTF-----TCP 115
 DB 189 kvqgdihprtpnpyllsalprerwnytlvldkdaefclrlhpsqplffewrptgtgr 248
 QY 116 FGTFAVRMPFGICNAPATQRCMAIFSNFCENIVEV---FMDDPSYVGSFDDCLSN 171
 DB 249 tggltwrlrpgfknspflfdeahlrldanfrlqhpvtllgyvddlllagatkqdcleg 308
 QY 172 LDRVLOKCKDTNLVNEKCHFMWNEGIVLGHKISEKGEVDAKAKYDAIDKMPYPTDIK 231
 DB 309 tkaillelsdlygrasakkaqicrrevlylgyrlgqgtwltearkktvwqipaptakg 368
 QY 232 IRSFLGHGFGFYRRFIKDF 249
 DB 369 vreflgtagfcrllwipgf 386

RESULT 12

AAW39272
 ID AAW39272 standard; Protein; 1194 AA.

XX AAW39272;

XX 19-MAY-1998 (first entry)

XX Porcine retrovirus POL protein.

XX Porcine retrovirus; POL protein; ENV protein; GAG protein;

XX vaccine; diagnosis; xenotransplantation; prophylactic; therapeutic.

XX Porcine retrovirus.

XX WO9740167-A1.

XX 30-OCT-1997.

XX 18-APR-1997; 97WO-GB01087.

XX 10-FEB-1997; 97GB-0002668.
 PR 19-APR-1996; 96GB-0008164.
 XX (IMUT-) IMUTRAN LTD.
 PA (COONE-) O-ONE BIOTECH LTD.
 XX Galbraith DN, Haworth C, Lees GM, Smith KT;
 DR WPI: 1997-535851/49.
 XX N-PSDB: AAV09700.
 PT Polynucleotide encoding porcine retrovirus expression product -
 PT useful to develop products for use in vaccines, diagnosis and
 PT xeno-transplantation
 XX
 PS Claim 6; Fig 3; 69pp; English.
 CC This sequence represents the porcine retrovirus (PoEV) polymerase (POL)
 CC protein. This protein and other porcine retroviral proteins e.g. the
 CC virion core (GAG) and envelope (ENV) proteins can be used to develop
 CC viral vaccines, antisense nucleic acids, ribozymes and other antiviral
 CC agents. They can also be used in xeno-transplantation technology and as
 CC diagnostic tools.
 XX
 SQ Sequence 1194 AA;

Query Match 14.1%; Score 190.5; DB 18; Length 1194;
 Best Local Similarity 25.6%; Pred. No. 2.1e-11;
 Matches 66; Conservative 42; Mismatches 121; Indels 29; Gaps 6;

QY 3 VKLLDEGIIVAHSDWSPVHSYPKKGITVVPNDKDELIPQRIITGYRMVIDFRKLN 62
 Db 196 Vgrllggllvpr-qspwntpllpvrkpg-----tnd-----yypvqdlrevn 237
 QY 63 KATRKDHPLPFIIDHMERL-SKLTNFCFLDGYSSFSQIPVAQSDQKTFP-----TCP 115
 Db 238 krvgdlhptvnpnyllalslaperntwylvldkdaflcrlhpsqplfafewrdpgtgr 297
 QY 116 GGTFAVRRMPFGLCNAPATFORCMWAIFFSNFCENIVEV----FMDDFSVYGSSPDDCLSN 171
 Db 298 tggltwtclpvgfknsplfdeahrdlanfrlqpyvtllqyvddlllagatkqdcleg 357
 QY 172 LDRVLQRCCKDTNLVNGEKCFFWNEGIVLGHKISERGIEVDKAKVDAIDKMPYPTDIK 231
 Db 358 tkallllsdlgyrasakqagircrevlylgyslrgqgrwltcarkkvvgjlapptakg 417
 QY 232 IRSFLGHGGFYRRFIKDF 249
 Db 418 vreflgtagfcrllwipgf 435

RESULT 13
 AAM32097
 ID AAM32097 standard; Protein; 1145 AA.
 XX
 AC AAM32097;
 XX
 DT 09-FEB-1998 (first entry)
 XX
 DE Miniature swine retrovirus POL protein.
 XX
 KM Retrovirus; porcine; POL protein; xenotransplantation; infectious;
 KM provirus; organ transplant; donor; activated virus; PCR.
 XX
 OS Porcine retrovirus.
 XX
 FH Key Location/Qualifiers
 FT Protein 1..1145
 XX /label= POL_protein
 XX
 PN MO9721836-A1.

XX 19-JUN-1997.
 PD
 XX 13-DEC-1996; 96MO-US19680.
 PF
 XX 14-DEC-1995; 95US-0572645.
 PR
 XX (GEHO) GEN HOSPITAL CORP.
 PA
 XX Fishman JA;
 PI
 XX WPI: 1997-332804/30.
 DR N-PSDB: AAT74884.
 PT New nucleic acid from porcine retroviruses - used for detecting
 PT viruses in transplant or other tissue and for assessing risk of
 PT transmitting infection to graft recipient
 XX
 PS Claim 22; Fig 3; 128pp; English.

This is a porcine retrovirus from miniature swine containing the coding
 region for a putative viral POL protein. This sequence and PCR fragments
 generated from the sequence (see AAT74812-T74882) could be used to
 screen organs for porcine retroviruses prior to xenotransplantation.
 CC Transplantation can increase the likelihood of retroviral activation if
 CC intact and infectious proviruses are present. The porcine retroviral
 CC sequence can be used to generate probes to determine the level (e.g.
 CC copy number) of intact (i.e. potentially replicating) porcine provirus
 CC sequences in a strain of xenograft transplantation donors. It can be
 CC used to detect mutations, genetic lesions or viral recombinants and
 CC also to determine the histological localisation of activated retrovirus.
 CC using Polymerase Chain Reaction DNA Quantitation (PQD) on blood
 CC mononuclear cells, infectivity titration and susceptibility testing can
 CC be performed. Ultimately animal donors without intact porcine retroviral
 CC sequences or a lower copy number of viral elements could be selected.

Sequence 1145 AA;

Query Match 14.1%; Score 189.5; DB 18; Length 1145;
 Best Local Similarity 25.6%; Pred. No. 2.5e-11;
 Matches 66; Conservative 44; Mismatches 119; Indels 29; Gaps 6;

QY 3 VKLLDEGIIVAHSDWSPVHSYPKKGITVVPNDKDELIPQRIITGYRMVIDFRKLN 62
 Db 147 Vgrllggllvpr-qspwntpllpvrkpg-----tnd-----yypvqdlrevn 188
 QY 63 KATRKDHPLPFIIDHMERL-SKLTNFCFLDGYSSFSQIPVAQSDQKTFP-----TCP 115
 Db 189 krvgdlhptvnpnyllalslaperntwylvldkdaflcrlhpsqplfafewrdpgtgr 248
 QY 116 GGTFAVRRMPFGLCNAPATFORCMWAIFFSNFCENIVEV----FMDDFSVYGSSPDDCLSN 171
 Db 249 tggltwtclpvgfknsplfdeahrdlanfrlqpyvtllqyvddlllagatkqdcleg 308
 QY 172 LDRVLQRCCKDTNLVNGEKCFFWNEGIVLGHKISERGIEVDKAKVDAIDKMPYPTDIK 231
 Db 309 tkallllsdlgyrasakqagircrevlylgyslrgqgrwltcarkkvvgjlapptakg 368
 QY 232 IRSFLGHGGFYRRFIKDF 249
 Db 369 mreflgtagfcrllwipgf 386

RESULT 14
 AAB73286
 ID AAB73286 standard; Protein; 1145 AA.
 XX
 AC AAB73286;
 XX
 DT 23-MAY-2001 (first entry)
 XX
 DE Retroviral protein #2 found in miniature swine.

```

XX KM Retrovirus; graft transplantation; xenotransplantation; miniature swine.
XX OS Unidentified.
XX PN US6190861-B1.
XX PD 20-FEB-2001.
XX PF 13-DEC-1996; 96US-0766528.
XX PR 14-DEC-1995; 95US-0572645.
XX PA (GEHO ) GEN HOSPITAL CORP.
XX PI Fishman JA;
XX DR MPI: 2001-256211/26.
XX N-PSDB; AAF77727.
XX PT Assessing risk of endogenous retroviruses in clinical practice and in
XX PT xenotransplantation, comprises using probe sequences derived from swine
XX PT or miniature swine retroviral genome -
XX PS Disclosure: Fig 3; 127pp; English.
XX CC The present invention relates to a method for screening a cell or tissue
XX CC for the presence or expression of a retrovirus (RV), comprising
XX CC contacting a target nucleic acid from the cell or tissue with a second
XX CC nucleic acid from the present invention (e.g. AAF77727 or a fragment
XX CC thereof). The method is useful for RV detection and to assess graft
XX CC transplantation risk. Screening of animals allows the elimination of
XX CC donors with active replication of known viruses. Inactive proviruses can
XX CC be detected and inactivated, allowing identification and elimination of
XX CC potential human pathogens derived from swine in a manner not possible in
XX CC the outbred human organ donor population and is important to the
XX CC development of human xenotransplantation.
XX SQ Sequence 1145 AA;

Query Match 14.1%; Score 189.5; DB 22; Length 1145;
Best Local Similarity 25.6%; Pred. No. 2.5e-11;
Matches 66; Conservative 44; Mismatches 119; Indels 29; Gaps 6;

OY 3 VVKLDEGIIVHANSWVSPVHSPKKGITVVPNDKDELIPQRIITGYRMYIDFRKLN 62
DB 147 VGLLGGILLPV-GSPWMLPILPVKPG-----LND-----YRPVQDLREVH 188
OY 63 KATRKHYLPFLIDHMLERL-SKLTHFCFLDGYSSFSQIPVAOSDQEKTF-----TCP 115
DB 189 KRYQDHPVTPNYPYLLICLPQTSWYLVLDKDAFFCLIRHPSQPLFAFEWRDPGTCR 248
OY 116 FGFFAVRMPFGICNAPATFORCMAIFSNFCENIVEV---FMDFSVYSSFPDCLSN 171
DB 249 TGGTITRIPGFKNSPTLFDALHNDANFRIGPVQLLYVDDLLAAGATKDCLEG 308
OY 172 LDRVLORCKDTNLVNGECKHEWNEGIVLGHKISERGLEVDKAKVDAIDMKRPYTDIKG 231
DB 309 TKALLLELSDLYGRASAKKAGICREVTLYGSLRDGQVLTQARKTKTVGJIPAPTAKQ 368
OY 232 IRSEFGHGFEYRRTIKDF 249
DB 369 MREFLGTAGTCITWLPGF 386

RESULT 15
AAR05898
ID AAR05898 standard; protein; 1784 AA.
AC AAR05898;
XX
XX 29-NOV-1990 (first entry)

```

```

XX DE Gene product of first ORF of virus capable of inducing viraemia in
XX DE feline species encoded by clone 61E.
XX KW Viraemia; Leukaemia; FeLV-A; clone 61C; clone 61E; ds.
XX OS Feline leukaemia virus-A.
XX PN EP377842-A.
XX PD 18-JUL-1990.
XX PF 12-DEC-1989; 89EP-0122964.
XX PR 13-DEC-1988; 88US-0284139.
XX PA (HARD ) HARVARD UNIV.
XX PA (COLS ) COLARADO STATE UNIV.
XX PI Hoover EA, Mullins JT;
XX DR MPI: 1990-218326/29.
XX N-PSDB; AAO05252.
XX PT DNA encoding virus-A sub: type - producing AIDS type disease in
XX PT cats, used to test drugs and vaccines.
XX PS Claim 6; Fig 1.1-1.7; 23pp; English.
XX CC Variant of viral genome induces immunodeficiency in cats similar to
XX CC AIDS. May be used in research, especially in testing drugs and
XX CC vaccines against viraemia and and feline leukaemia viruses.
XX SQ Sequence 1784 AA;

Query Match 13.0%; Score 174.5; DB 11; Length 1784;
Best Local Similarity 23.2%; Pred. No. 2e-09;
Matches 60; Conservative 50; Mismatches 118; Indels 31; Gaps 7;

OY 5 KLDDEGIIVHANSWVSPVHSPKKGITVVPNDKDELIPQRIITGYRMYIDFRKLNKA 64
DB 780 RMLDQGL-KPCGSPWMLPILPVK-----PGLKD-----YRPVQDLREVHNR 821
OY 65 TRKHYPFLPFLIDHMLERLTH--FCFLDGYSSFSQIPVAOSDQEKTFCTCP-----F 116
DB 822 VEDLHPVTPNYPYLLICLP-SHPWYLVLDKDAFFCLIRHPSQPLFAFEWRDPGTCR 880
OY 117 GTFFAVRMPFGICNAPATFORCMAIFSNFCEN---IVEVFMDFSVYSSFPDCLSNL 172
DB 881 GQLTWRIPGFKNSPTLFDALHNSDLADFVRYPALVLLQYVDDLLAATRECELEG 940
OY 173 DRYLORCKDTNLVNGECKHEWNEGIVLGHKISERGLEVDKAKVDAIDMKRPYTDIKGI 232
DB 941 KALLELTGNKGRASAKKAGICREVTLYGSLKDGQVLTQARKTKAELISIPKXBPYV 1000
OY 233 RSEFGHGFEYRRTIKDF 251
DB 1001 REFLGTAGTCITWLPGF 1019

Search completed: April 1, 2002, 22:42:41
Job time: 4388 sec

```

...

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-724 <STO>
A:Cross-references: GB:NC_001268; NID:g7267363; PDB:1CAB8136.1; GSPDB:GN00140
C:Genetics:
A:Gene: AT4g07860
A:Map position: 4

Query Match	62.48;	Score 841;	DB 2;	Length 724;
Best Local Similarity	66.88;	Pred. No. 3.9e-6;		
Matches 155;	Conservative 32;	Mismatches 45;	Indels 0;	Gaps 0;

QY	1	K E V A L D E G I J T Y H A N S D W M S P H S V P K R C G I T U V P N D K R E L I P O R I T T G Y R M V I D E R K	60
D b	367	K E I L K L D A G I T Y I P S D S T W F P A P H C V P K G G M V V K N E K E L L I P T R I T T S H R A V C I D Y R K	422
QY	61	L N A K T R K D H Y L P E I D H L E R I S K L T H P C E F J D G S S F S Q I P V A O S D D E K T F T G P C F G F A	120
D b	427	L N A S R K D H P L P P T N O M L E G L A N H L V N C F L D G T S G F Q I P I H N D D E K T F T P R Y G T F A	486
QY	121	Y R R M P F G L C N A P A T F O R C M A I F S N F C E N I V E P M D F S V Y G S S F D C L S M L D I V L O R K	180
D b	487	Y K R M P F G L C N A P T F O R C M T S I F S D L I E K M E V A M D F S V Y G P S F S C L L N G V L R K M E	546
QY	181	D T N L V L G E C K F M V N G C I V L G H I S E R G T F V D A K A Y D A I D K M Y P P I D I G I	232
D b	547	E T N L V L M W E K C Y F M V K E G I V L G H I S K G I E V D E K T K V M Q I O P P T V A D I	598

RESULT 3
D84513
C: Probable retroelement pol polypeptide [imported] - Arabidopsis thaliana
C: Species: Arabidopsis thaliana (mouse-ear cress)
C: Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C: Accession: D84513
R: Lin, X., Kaul, S., Rounsley, S.D., Shea, T.P., Benito, M.I., Town, C.D., Fujii
M., Ko, H., Moffatt, K.S., Cronin, L.A., Shen, M., Vanden, S.E., Umayan, L., Ta
eus, D., Nierman, W.C., Williams, A.C., White, O., Eisen, J.A., Salzberg, S.L., Fraser, C.M., V
Nature 402, 761-768, 1999
A: Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana
A: Reference number: A84420; MIMD:20083487
A: Accession number: A84420; MIMD:20083487

Query Match	61.3%,	Score 826;	DB 2;	Length 841;
Best Local Similarity	62.2%,	Pred. No. 1e-65;		
Matches 156;	Conservative 30;	Mismatches 37;	Indels 28;	Gaps 2

00Y	1	KEVVALLDGGILVHVAHSDWSPVSHSPKKGGITVVPNDKDEILPORIINGYRAVIDERK	60
00Y	392	KEIMKLLDGGIIPYISDSTWSPVHVPPKGGVIVIKNENNELIPRTVYGHMCCIDYRK	451
00Y	61	LKATKTHKHYPPLFDIMHLERLSKLTHFCTLDGYSSEFSQIPVAOSDOEKTTFCPTGTA	120
00Y	452	LNSATKRRDNFPLSEIDOMERLSNOCPYCFLLDYLGFQIILHPDDOEKTTFCPTGTA	511
00Y	121	YRMRPGLCNAPATFQRCMMATFSPNCCENTVEVFMDFSVYGGSPDCCSLNDRVLORCK	180
00Y	512	YRMRPGLCNAPATFQRCMMATFSPNCCENTVEVFMDFSVYGGSPDCCSLNDRVLORCK	556
00Y	181	DTNLVLNGEKGCFMWNVEGIVLGHIKISEERGIEVDKAKVADAIKMPYPTDIKIGISPLGHG	240
00Y	557	DKNLVLNMEKSHPMYVADGIVLGHIKISEKGEVENDRAKIE-----IKRFLGHAG	603
00Y	241	FYRRFLKDETK	251

Db 604 FYRRFIKDFSK 614

RESULT 4
T12085

Reverse transcriptase homolog - fava bean (fragment)
C:Species: *Vicia faba* (fava bean)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #ext_change 03-Nov-2000
C:Accession: T12085
R:Kinoshita, T.; Wada, H.; Masaaki, I.; Shimazaki, K.
submitted to the EMBL Data Library, September 1997
A:Description: Retrotransposon-like cDNAs from guard cell protoplasts in *Vicia faba*.
A:Reference number: z17406
A:Accession: T12085
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-407 <KIN>
A:Cross-references: EMBL:AB007466; NID:d1170509; PID:d1023656
A:Experimental source: guard cell protoplasts
C:Superfamily: pol polypeptide

Query Match	57.5%	Score 774;	DB 2;	Length 407;
Best Local Similarity	70.9%;	Pred. No. 1.9e-61;		
Matches 139;	Conservative 29;	Mismatches 28;	Indels 0;	Gaps 0

[illegible]

RESULT 5
B96492
probable polyprotein, 77260-80472 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: B96492
R:Thelloglis, A.; Ecker, J.R.; Palm, C.J.; Federpsiel, N.A.; Kaul, S.; White, O.; Alonso
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
ansen, N.F.; Hughes, B.; Hultzar, L.
Nature 408, 816-820, 2000
C:A:; Li, J.H.; Li, Y.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzita
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shim, P.; Southwick, A.M.; Sun, H.; Tallo
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: B96492
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-884 <STO>
A:Cross-references: GB:AE005173; NID:g10092524; PIDN:AA612922.1; GSPDB:GN00141
C:Genetics:
A:Gene: T4T21.10
;Map position: 1

Query Match	47.8%;	Score 644;	DB 2;	Length 884;
Best Local Similarity	58.1%;	Pred. No. 1.9e-49;		

Db 475 SNFKIQMDKSEFLKLETAVLGHIIISRDGIKPNPKISAIOKYLIPTKPEIKQFLGLGY 534
 QY 242 YRRFIKDFTK 251
 ||:|||||:
 Db 535 YRKFIIPDFAR 544

RESULT 9
 534639
 pol protein - fruit fly (Drosophila ananassae) transposon Tom (fragment)
 C:Species: Drosophila ananassae
 C>Date: 22-Nov-1993 #sequence_revision 01-Dec-1995 #text_change 26-Aug-1999
 C:Accession: S34639
 R:Panda, S.; Corces, V.G.
 Submitted to the EMBL Data Library, July 1993
 A:Description: Structural and functional analysis of a Drosophila retrotransposon that
 A:Reference number: S34638
 A:Accession: S34639
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1040 <TAN>
 A:Cross-references: EMBL:224451; NID:9394703; PIDN:CAA80824.1; PID:9394705
 C:Genetics:
 A:Gene: FLYBase:Dana/Tom
 A:Cross-references: FLYBase:Fbgn0004357
 A:Mobile element: retrotransposon Tom
 C:Superfamily: pol polyprotein

Query Match 32.4%; Score 436; DB 2; Length 1040;
 Best Local Similarity 37.9%; Pred. No. 8.6e-31;
 Matches 94; Conservative 49; Mismatches 91; Indels 14; Gaps 3;
 QY 2 EVVKLLDEGIIVHANSWVSPVHSPKGGITVVPNDKDELIPORITGVRMVIDPERKL 61
 ||:|||||:
 Db 186 QVQEMLEGGIT-RESNSPYNSPTWVVPK-----PDASG-----AKRVVIDYRKL 231
 ||:|||||:
 QY 62 NKATRKDHPLPFIIDHMLERLSKLTGHCFLDGSSFSQIPVAOSDQETTTCPREGTAY 121
 ||:|||||:
 Db 232 NETTIDRPRIPMDIELGKLCQYFTTIDLAGRFHQLMESESIOKTAFTSKRGHYEX 291
 ||:|||||:
 QY 122 RMPFGLCNAPATFORCMAAIFSNFCENIVEFMDDFSVGSSFDCLSNLDVLRCKD 181
 ||:|||||:
 Db 292 VMPFGLCNAPATFORCMAAIFSNFCENIVEFMDDFSVGSSFDCLSNLDVLRCKD 351
 ||:|||||:
 QY 182 TNLVNGEKCHPMVNEGIVLGHKISERGIENVDAKAYDAIDKMPYPTDINGIRSLGHGCF 241
 ||:|||||:
 Db 352 SNLKLQDLKCEFLKKEATFGLGHVTPDGIRKPNPLKVEALASYPITPKVEIRAFILGTMGY 411
 ||:|||||:
 QY 242 YRRFIKDF 249
 ||:|||||:
 Db 412 YRKFIIPSY 419

RESULT 10
 534639
 retrovirus-related pol polyprotein - fruit fly (Drosophila melanogaster) retrotransposon
 N:Alternate names: reverse transcriptase
 C:Species: Drosophila melanogaster
 C>Date: 04-Dec-1986 #sequence_revision 04-Dec-1986 #text_change 16-Jun-2000
 C:Accession: A03971
 R:Salgo, K.; Kogut, W.; Matsuo, Y.; Inouye, S.; Yoshio, K.; Yuki, S.
 Nature 312, 659-661, 1984
 A:Title: Identification of the coding sequence for a reverse transcriptase-like enzyme
 A:Reference number: A93349; MUID:85061628
 A:Accession: A03971
 A:Molecule type: DNA
 A:Residues: 1-1058 <SAI>
 A:Cross-references: GB:X01472; GB:J01060; GB:J01061; NID:98142; PIDN:CAA25702.1; PID:913
 C:Genetics:
 A:Gene: FLYBase:17.6
 A:Cross-references: FLYBase:Fbgn0000004
 C:Superfamily: pol polyprotein

C:Keywords: polyprotein; reverse transcriptase

Query Match 31.7%; Score 427; DB 1; Length 1058;
 Best Local Similarity 37.9%; Pred. No. 5.6e-30;
 Matches 94; Conservative 47; Mismatches 93; Indels 14; Gaps 3;
 QY 2 EVVKLLDEGIIVHANSWVSPVHSPKGGITVVPNDKDELIPORITGVRMVIDPERKL 61
 ||:|||||:
 Db 226 QIOMDLNQGII-RTNSPYNSPIMVVPK-----QDASGKAK-----PRIVIDYRKL 271
 ||:|||||:
 QY 62 NKATRKDHPLPFIIDHMLERLSKLTGHCFLDGSSFSQIPVAOSDQETTTCPREGTAY 121
 ||:|||||:
 Db 272 NETTIDRPRIPMDIELGKLCQYFTTIDLAGRFHQLMESESIOKTAFTSKRGHYEX 331
 ||:|||||:
 QY 122 RMPFGLCNAPATFORCMAAIFSNFCENIVEFMDDFSVGSSFDCLSNLDVLRCKD 181
 ||:|||||:
 Db 332 RMPFGLCNAPATFORCMAAIFSNFCENIVEFMDDFSVGSSFDCLSNLDVLRCKD 391
 ||:|||||:
 QY 182 TNLVNGEKCHPMVNEGIVLGHKISERGIENVDAKAYDAIDKMPYPTDINGIRSLGHGCF 241
 ||:|||||:
 Db 392 ANLKLQDLKCEFLKKEATFGLGHVTPDGIRKPNPLKVEALASYPITPKVEIRAFILGTMGY 451
 ||:|||||:
 QY 242 YRRFIKDF 249
 ||:|||||:
 Db 452 YRKFIIPNF 459

RESULT 11
 T01842
 hypothetical protein F9D12.11 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C>Date: 26-Feb-1999 #sequence_revision 26-Feb-1999 #text_change 24-Mar-1999
 C:Accession: T01842
 R:Murray, J.; Langston, Y.; Ahrens, C.
 Submitted to the EMBL Data Library, July 1998
 A:Description: The sequence of Arabidopsis thaliana F9D12.
 A:Reference number: Z14444
 A:Accession: T01842
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1522 <MUR>
 A:Cross-references: EMBL:AF077407; NID:93319339; PID:93319351
 A:Experimental source: cultivar Columbia
 C:Genetics:
 A:Map position: 4
 A:Introns: 180/2; 949/1; 1048/3
 A:Note: F9D12.11

Query Match 31.7%; Score 427; DB 2; Length 1322;
 Best Local Similarity 39.2%; Pred. No. 7.3e-30;
 Matches 91; Conservative 37; Mismatches 86; Indels 18; Gaps 1;
 QY 18 SDWSPVHSPVPPKGGITVVPNDKDELIPORITGVRMVIDPERKLKATRKDHPLPFIIDH 77
 ||:|||||:
 Db 557 SPWRAPMLPMPKRRKG-----SFLCIDYGLNGVYTKNKPPLPRIDE 598
 ||:|||||:
 QY 78 MIERLSKLTGHCFLDGSSFSQIPVAOSDQETTTCPREGTAYRMPFGLCNAPATFOR 137
 ||:|||||:
 Db 599 LLDQRLGATCFSKIDTSDYHOIPAEADVKTARTRTGHEFEVMPFGLCNAPAAFMAR 658
 ||:|||||:
 QY 138 CMAAIFSNFCENIVEFMDDFSVGSSFDCLSNLDVLRCKDYNLVNGEKCHPMVNE 197
 ||:|||||:
 Db 659 LMNSVFOEFLDEFVLIFIDILLVYSPEEHVHLARRKYEKRLAKSLKSCGFQORE 718
 ||:|||||:
 QY 198 GIVGKHSERGIENVDAKAYDAIDKMPYPTDINGIRSLGHGCFYRRFIKDF 249
 ||:|||||:
 Db 719 MGLFISHIVSAGVSVDPKIEAIRDMRPRTNAVEIRSLFGLAGYRRRVKGF 770

RESULT 12
 A84460

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 1, 2002, 22:42:43 ; Search time 67.95 Seconds
(without alignments)
135,436 Million cell updates/sec

Title: US-09-586-106-63

Perfect score: 1347
Sequence: 1 KEVVKLDEGIIVHAHSDW.....INSEFLGSGFYRRIFDKFK 251

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt_39:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	427	31.7	1058	POL3_DROME	P04323 drosophila
2	418	31.0	1059	POL2_DROME	P20823 drosophila
3	408.5	30.3	1237	POL4_DROME	P10394 drosophila
4	370.5	27.5	1035	POLX_DROME	P10401 drosophila
5	339	25.2	1333	RDPQ_SCHPO	P05654 schizosacch
6	292	21.7	1745	YL51_CAEEL	P34331 caenorhabdi
7	264.5	19.6	1268	YRD6_CAEEL	P09575 caenorhabdi
8	245.5	18.2	679	POL_CAVVC	P03555 cauliflower
9	245.5	18.2	679	POL_CAVVC	P02964 cauliflower
10	245.5	18.2	679	POL_CAVVC	P03554 cauliflower
11	242	18.0	142	RRPO_ORNBE	P31843 oenothera b
12	241.5	17.9	680	POL_CAVVC	P00862 cauliflower
13	238.5	17.7	674	POL_CAVVC	P03556 cauliflower
14	232.5	17.3	666	POL_FAVD	P09523 figwort mos
15	227.5	16.9	1675	POL_RTBVC	P27502 rice tungro
16	223	16.6	1886	POL_COVAV	P19199 commelina y
17	208	15.4	682	POL_SCKAV	P15629 soybean chl
18	192	14.3	659	POL_CERY	P05400 carnation e
19	191.5	14.2	1165	POL_GALV	P21414 gibdon ape
20	190.5	14.1	1046	POL_FENV1	P31792 feline endo
21	186.5	13.8	1189	POL_BAEVM	P10272 baboon endo
22	168.5	12.5	1161	POL_SFV1	P23074 simian foam
23	164.5	12.2	1204	POL_MLVFV	P26809 friend muri
24	164.5	12.2	1204	POL_MLVFV	P27401 friend muri
25	163.5	12.1	1157	POL_SFV3L	P26810 friend muri
26	159.5	11.8	1204	POL_MLVFV	P14350 human spuna
27	159	11.8	886	POL_FOAMV	P03355 moloney mur
28	158.5	11.8	1199	POL_MLVAV	P03356 akv murine
29	151.5	10.5	1196	POL_MLVAV	P11227 radiation m
30	141.5	9.9	1056	POL_BIV06	P19560 bovine immu
31	134	9.9	1056	POL_BIV27	P19561 bovine immu
32	134	9.9	1056	POL_BIV27	P19561 bovine immu
33	133.5	9.9	1046	POL_SIVAG	P27980 simian immu

34	129.5	9.6	863	1	POL_IPHA	P04026 hamster int
35	127.5	9.5	852	1	POL_BIVAU	P25059 bovine leuk
36	125.5	9.3	852	1	POL_BIV2	P03361 bovine leuk
37	124.5	9.2	1035	1	POL_HV2NZ	P05962 human immu
38	124.5	9.2	1036	1	POL_HV2RO	P04584 human immu
39	123.5	9.2	1019	1	POL_SIVS4	P12502 simian immu
40	123.5	9.2	1035	1	POL_HV2SR	P12451 human immu
41	123.5	9.2	1035	1	POL_HV2SB	P12451 human immu
42	120.5	8.9	1055	1	POL_HV2ST	P20876 human immu
43	120.5	8.9	1022	1	POL_SIVSP	P19505 simian immu
44	119.5	8.9	1142	1	POL_HV2BE	P18096 human immu
45	118.5	8.8	1009	1	POL_SIVGB	P22382 simian immu

ALIGNMENTS

RESULT 1

ID	POL3_DROME	STANDARD:	PRT:	1058 AA.
AC	P04323	20-MAR-1987 (Rel. 04, Created)		
DT	20-MAR-1987 (Rel. 04, Last sequence update)			
DT	20-AUG-2001 (Rel. 40, Last annotation update)			
DE	RETROVIRUS-RELATED POL. POLYPROTEIN FROM TRANSPOSON 17.6 [CONTAINS: PROTEASE (EC 3.4.23.-); REVERSE TRANSCRIPTASE (EC 2.7.7.49);			
DE	ENDONUCLEASE].			
GN	POL.			
OS	Drosophila melanogaster (Fruit fly).			
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;			
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
OX	Ephydroidea; Drosophilidae; Drosophila.			
NCBI_TaxID=7227;				
NP	SEQUENCE FROM N.A.			
RP	MEDLINE=85061628; PubMed=6209583;			
RA	Salgo K., Kugimiyu W., Matsuo Y., Inouye S., Yoshioke K., Yuki S.;			
RT	*Identification of the coding sequence for a reverse transcriptase-like enzyme in a transposable genetic element in Drosophila melanogaster.*			
RT	Nature 312:659-661(1984).			
RL	-I- MISCELLANEOUS: THE OPEN READING FRAME IS LOCATED IN A COPIN-LIKE			
CC	-I- TRANSPOSABLE ELEMENT CALLED 17.6.			
CC	-I- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY U22.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	EMBL: X01472; CA25702.1; -			
DR	PIR: A03971; GNPEF17.			
DR	HSSP: P04585; IRTI.			
DR	MEROPS: A02.052; -			
DR	FLYBASE: FBgn0000004; 17.6.			
DR	InterPro: IPR001995; Asp_prot.retrov.			
DR	InterPro: IPR00477; RVISE.			
DR	InterPro: IPR001584; Rve.			
DR	Pfam: PF00665; rve; 1.			
DR	Pfam: PF00077; rvp; 1.			
DR	Pfam: PF00078; rvt; 1.			
DR	PROSITE: PS00141; ASP_PROTEASE; 1.			
KW	Hydrolase; Aspartyl protease; RNA-directed DNA polymerase;			
KW	Endonuclease; Transferrase; Polypeptide; Transposable element.			
FT	ACT_SITE 30			
FT	ACT_SITE 30			
SQ	SEQUENCE 1058 AA; 122697 MW; C893F5C4A7E1F091 CRC64;			

Query Match 31.7%; Score 427; DB 1; Length 1058;

Best Local Similarity 37.9%, Pred. No. 5.9e-32;
Matches 94; Conservative 47; Mismatches 93; Indels 14; Gaps 3;

QY 2 EVKLLDGIILYVHAHSMVSPVHSGGTTVVPNKDKDELIPRIITGYMVDKFL 61
DQ 226 QIOQMLNGIIL-RSNSNSPIWVPRK-----QDSGKOK---FRIVIDYRKL 271
QY 62 NKATRKDHVPLPFDHMLERLSTLHFCFLDGYSSFSQIPVAQSDQEKTFCTPCPTAY 121
DQ 272 NEITVGRHPIPNMDELIGKLGKQYFTTIDLAKGFHQIEMDEESISTATSKGHYEX 331
QY 122 RRMFGCLNAPATFORCMMALFNSNCEIVEFMDDESFGSSPDDCLSNDRVLQKRD 181
DQ 332 LRMPFGLKNAPATFORCMMNLRLPLNKHCLVYDDIIFSVSLHNSIQIOLVETKIALD 391
QY 182 TNVLNGEKGHPVNEGIVLGHKISERGIEVDKAKVDAIDKMPPTDIKGRSFLGHGCF 241
DQ 392 ANLKIQLDKCEFLKKEANFLGHIVTPDGIRKPNPEIKAIQYPIPTKKEIKAFGLTGY 451
QY 242 YRRFIKDF 249
DQ 452 YRRFIKDF 459

RESULT 2

POL2_DROME STANDARD; PRT; 1059 AA.

AC P20825;
DT 01-FEB-1991 (Rel. 17, Created)
DT 20-AUG-2001 (Rel. 17, Last sequence update)
DE RETROVIRUS-RELATED POL. POLYPROTEIN FROM TRANSPOSON 297 [CONTAINS:
DE ENDONUCLEASE].
GN POL.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_Taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86108354; PubMed=2417839;
RA Inoue S., Yuki S., Saigo K.;
RT "Complete nucleotide sequence and genome organization of a Drosophila
transposable genetic element, 297.";
RL Eur. J. Biochem. 154:417-425(1986).
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY U22.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch)
EMBL: X03431; CA557796.1; ALT_SEQ.
DR PIR: B24872; B24872.
DR HSSP: P04585; IRTI.
DR FlyBase: FBgn0000005; 297.
DR InterPro: IPR001995; Asp_prot_retrov.
DR InterPro: IPR001969; Asp_protease.
DR InterPro: IPR000477; RVTse.
DR InterPro: IPR001584; Rve.
DR Pfam: PF00665; Irv; 1.
DR Pfam: PF00077; Irv; 1.
DR Pfam: PF00078; Rvt; 1.
DR PROSITE: PS00141; Asp_PROTEASE; 1.
KM Hydroxylase; Aspartyl protease; RNA-directed DNA polymerase;
KW Endonuclease; Transferrase; Polypeptide; Transposable element.
FT ACN_SITE 30
FT ACN_SITE 30
SEQUENCE 1059 AA; 123310 MW; 3905CF38E914173D CRC64;

Query Match

Best Local Similarity 31.0%, Score 418; DB 1; Length 1059;
Matches 94; Conservative 44; Mismatches 96; Indels 14; Gaps 3;

QY 2 EVKLLDGIILYVHAHSMVSPVHSGGTTVVPNKDKDELIPRIITGYMVDKFL 61
DQ 225 QVOEMLNGIIL-RESNSPNSPTWVPRK-----PDASG-----ANKYRVVIDYRKL 270
QY 62 NKATRKDHVPLPFDHMLERLSTLHFCFLDGYSSFSQIPVAQSDQEKTFCTPCPTAY 121
DQ 271 NEITVGRHPIPNMDELIGKLGKQYFTTIDLAKGFHQIEMDEESISTATSKGHYEX 330
QY 122 RRMFGCLNAPATFORCMMALFNSNCEIVEFMDDESFGSSPDDCLSNDRVLQKRD 181
DQ 331 LRMPFGLKNAPATFORCMMNLRLPLNKHCLVYDDIIFSVSLHNSIQIOLVETKIALD 390
QY 182 TNVLNGEKGHPVNEGIVLGHKISERGIEVDKAKVDAIDKMPPTDIKGRSFLGHGCF 241
DQ 391 ANLKIQLDKCEFLKKEANFLGHIVTPDGIRKPNPEIKAIQYPIPTKKEIKAFGLTGY 450
QY 242 YRRFIKDF 249
DQ 451 YRRFIKDF 458

RESULT 3

POL4_DROME STANDARD; PRT; 1237 AA.

AC P10394;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE RETROVIRUS-RELATED POL. POLYPROTEIN FROM TRANSPOSON 412 [CONTAINS:
DE PROTEASE (EC 3.4.23.-); REVERSE TRANSCRIPTASE (EC 2.7.7.49);
DE ENDONUCLEASE].
GN POL.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_Taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86274717; PubMed=2426108;
RA Yuki S., Inoue S., Ishimaru S., Saigo K.;
RT "Nucleotide sequence characterization of a Drosophila
retrotransposon, 412.";
RL Eur. J. Biochem. 158:403-410(1986).
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY U22.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch)
EMBL: X04132; CAA27750.1; -
DR PIR: D29349; GNEP42.
DR HSSP: P03366; IHG3.
DR FlyBase: FBgn0000006; 412.
DR InterPro: IPR001995; Asp_prot_retrov.
DR InterPro: IPR001969; Asp_protease.
DR InterPro: IPR000477; RVTse.
DR InterPro: IPR001584; Rve.
DR Pfam: PF00665; Irv; 1.
DR Pfam: PF00077; Irv; 1.
DR Pfam: PF00078; Rvt; 1.
DR PROSITE: PS00141; Asp_PROTEASE; 1.
DR PROSITE: PS00175; Asp_PROT_RETROV; 1.


```

RT fission yeast.
RL Gene 131:135-139(1993).
RN (2)
RP SEQUENCE FROM N.A. (SPAC26A3.13C).
RC STRAIN-972;
RA McLean J., Harris D., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
RN (3)
RP SEQUENCE FROM N.A. (SPAC27E2.08).
RC STRAIN-972;
RA Murphy L., Harris D., Barrell B.G., Rajandream M.A., Wood V.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
RN (4)
RP SEQUENCE FROM N.A. (SPAC9.04).
RC STRAIN-972;
RA Weiler H., Duesterhoeft A., Lyne M.H., Rajandream M.A., Barrell B.G.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RN (5)
RP SEQUENCE FROM N.A. (SPAC9B6.02C).
RC STRAIN-972;
RA Wood V., Rajandream M.A., Barrell B.G., Volckaert G.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN (6)
RP SEQUENCE OF 120-1333 FROM N.A. (SPAC167.08).
RC STRAIN-972;
RA Rieger M., Wood V., Rajandream M.A., Barrell B.G.;
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: REGIONALITY, TO PROTEASE, REVERSE TRANSCRIPTASE,
CC RNASE H AND INTEGRASE FROM OTHER RETROTRANSPOSONS AND
CC RETROVIRUSES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L10324; AAA91215.1; -
DR EMBL: Z69240; CAA93236.1; -
DR EMBL: Z98978; CAB11682.1; -
DR EMBL: AL121764; CAB57422.1; -
DR EMBL: AL049769; CAB42363.1; -
DR EMBL: AL035248; CAB58169.1; -
DR MEROPS: A02.051; -
DR InterPro: IPR001969; Asp_protease.
DR InterPro: IPR000477; RVTse.
DR InterPro: IPR001584; Rve.
DR Pfam: PF00665; rve; 1.
DR Pfam: PF00078; rvt; 1.
DR PROSITE: PS00141; ASP_PROTEASE; 1.
KW transposable element.
FT VARIANT 1 400 TO CAPSID PROTEIN FROM RETROVIRUSES.
FT VARIANT 206 206 L->P (IN SPAC27E2.08).
SQ SEQUENCE 1333 AA; 154932 MW; 2F4540823BDC1DA CRC64;

```

Query Match 25.2%; Score 339; DB 1; Length 1333;
 Best Local Similarity 33.5%; Pred. No. 1,3e-23;
 Matches 75; Conservative 41; Mismatches 90; Indels 18; Gaps 1;

```

QY 23 PVLSVRKKGIVVNDKDELIPORITGRVVIDFRKLKATRKDHYPLPFIIDHMLERL 82
DB 451 PIVVKKKEGT-----LRMVVDYKPLKKYKPNITPLPIEQLAKI 492
QY 83 SKLTHCFIDGYSFSGIPVAOSDQKTTFTCPGTFAYRRMPFGICNATATQRCMAAI 142
DB 493 QSGTITTKLKLKSAHLIRKGDENHLARCGVFEYLYMPGISTADAHFYDTINTI 552
QY 143 FSNFCEIVFEMDFESVSGSFDCLSNDRVLYORCKDNTNLVKGKCFHFMVNEGIVLG 202
DB 553 LGRAKESHVYVCMDDLILHKSSEHKKHVKVDLQKLKNNLLINQAKCFHOSQVYF IG 612

```

```

QY 203 HKISRGIEVDKAKYDAIDKMPYPTDIKIRSEFLGHGTYRRFI 246
DB 613 YHISRGFTPCQENIDKYLQWKQPKNRKELROFLGSVNYLRFI 656

```

RESULT 6

YL51_CAEEL STANDARD; PRT; 1745 AA.

```

AC P34431.
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE HYPOTHETICAL PROTEIN F44E2.1 IN CHROMOSOME III (FRAGMENT).
GN F44E2.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=62339;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL NZ;
RA Anderson K.;
RL Submitted (SEP-1993) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO RETROVIRAL-TYPE PROTEASE/REVERSE TRANSCRIPTASE/
CC ENDONUCLEASE.
CC -----

```

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```

DR EMBL: L23646; AAA28035.1; -
DR HSSP: P04585; IRTI.
DR WormPep: F44E2.1; CE00542.
DR InterPro: IPR001969; Asp_protease.
DR InterPro: IPR000477; RVTse.
DR InterPro: IPR001584; Rve.
DR InterPro: IPR001878; Znf_CCHC.
DR Pfam: PF00665; rve; 1.
DR Pfam: PF00078; rvt; 1.
DR Pfam: PF00098; zf-CCHC; 1.
DR SMART: SM00343; znf-C2HC; 1.
DR PROSITE: PS00141; ASP_PROTEASE; 1.
KW Hypothetical protein; Hydrolase; Aspartyl protease;
KW Transferrase; RNA-directed DNA polymerase.
FT ACT_SITE 382 382
FT NON_TER 1745 1745
FT SEQUENCE 1745 AA; 198408 MW; 7C7457EECB19329 CRC64;

```

Query Match 21.7%; Score 292; DB 1; Length 1745;
 Best Local Similarity 29.1%; Pred. No. 4.2e-19;
 Matches 73; Conservative 40; Mismatches 100; Indels 38; Gaps 3;

```

QY 1 KEVYKLDEGIIVAVASDWSVPSVPRKGGITVAVNDKDELIPQILITGRVVIDFRK 60
DB 678 KTIQKMLNOKVT-RESKSPSSPVVLVKKRG-----SIRCIDYRK 718
QY 61 LNKATRKDHYPLPFIIDHMLERLSKLTHCFIDGYSFSGIPVAOSDQKTTFTCPGTF 120
DB 719 VNKVYKNNAPHLPIETLSTSLAGKILYTFDMTAGWQIPLDKSKETIAFAIGSLFE 778
QY 121 YRRMPFGICNAPATFGRCMAATFNSFCENIVFEMDFESVYSGSFDCLSNDRVLYORCK 180
DB 779 WNVLPFGLIVSPALFQGTMEIITGDLGVCIFYVDLLASKMEQHLDVKEALRRIR 838
QY 181 DTNLVNGEKCHPVRNNGIYVLGKHSRGIEVDKAKYDAIDKMPYPTDIKIRSEFLGHG 240
DB 839 KSGKILRASKCHIAKKKEVEYLGHKVTLLDGVETG-----SFLGLVG 879

```

QY 241 FYRRFKDFTK 251
 DB 880 YRRKFLNFAQ 890

RESULT 7

YRD6_CAEEL STANDARD; PRT; 1268 AA.

AC Q09575; 01-FEB-1996 (Rel. 33, Created)

DT 01-FEB-1996 (Rel. 33, Last sequence update)

DE 20-AUG-2001 (Rel. 40, Last annotation update)

HYPOHETICAL 143.5 KDA PROTEIN K02A2.6 IN CHROMOSOME II.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;

CC Rhabditidae; Peloderinae; Caenorhabditis.

NCBI_TaxID=6239;

CC -1- SIMILARITY: WEAK, TO RETROVIRUS-RELATED POLYPROTEIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC EMBL; 023171; AAC46707.1; -

DR HSSP; P04585; 1RT1.

DR WormRep; K02A2.6; CE02792.

DR InterPro; IPR000477; RVTse.

DR InterPro; IPR001584; Rve.

DR InterPro; IPR001878; Znf_CCHC.

DR Pfam; PF00665; rve; 1.

DR Pfam; PF00078; rvl; 1.

DR Pfam; PF00098; zf-CCHC; 1.

DR PRINTS; PR00939; C2HCZNRNBR.

DR SMART; SM00343; Znf_C2HC; 1.

KW Hypothetical protein.

SQ SEQUENCE 1268 AA; 143493 MW; AACAC4235AAE468 CRC64;

Query Match 19.6%; Score 264.5; DB 1; Length 1268;

Best Local Similarity 29.6%; Pred. No. 11e-16;

Matches 74; Conservative 44; Mismatches 107; Indels 25; Gaps 6;

QY 2 EVVKLDEGIIYVAHSDWSPVSHVSKKGIYVVPNDKDELIPORITNG-YRMVDFR- 59

DB 459 ELNRIDEMGIYVITAKMAPIVIAKKKG-----TGKIRVCADEFRC 500

QY 60 -KLNKATRKDHYPPLFDIHLERLSKLTHRCFLDGYSSSQIPVAOSDEKTTFTCPFGT 118

DB 501 SGLNALKDEHFLPTSEDFSRLL-KGVYSSQIDLDKAVLYQVLDLDEAKLAVINTHRI 559

QY 119 FAVRBPFGCLNAPATFORCMAIFSNFCENIVEVMDPSYVSSFDCLSLNDVLR 178

DB 560 FKYLRTFGLKPAFASQIKMDKMSGLTG--VAIVWDIIISASSIEHEKILRELFER 617

QY 179 CKDTNLVNGEKCHFWNNEGIVLGKHSERGIEVDRAKYDAIDKMPYPTDIDKIRSLGH 238

DB 618 FKREGFVNSAKCAFAQKQVTLPLG-VDEHGRPRDSKKTETALISMMAPTDQKQLASFLG 676

QY 239 GGEYRPFIKD 248

DB 677 ADWLIRMDQD 686

RESULT 8

POL_CAMVC STANDARD; PRT; 679 AA.

AC P03555; 21-JUL-1986 (Rel. 01, Created)

DT 21-JUL-1986 (Rel. 01, Last sequence update)

DE 20-AUG-2001 (Rel. 40, Last annotation update)

ENZYMOLOGICAL POLYPEPTIDE [CONTAINS: ASPARTIC PROTEINASE (EC 3.4.23.-);

DE ENDONUCLEASE; REVERSE TRANSCRIPTASE (EC 2.7.7.49)].

OS Cauliflower mosaic virus (strain CM-1841) (CaMV).

OC Viruses; Retroviruses; Caulimovirus.

NCBI_TaxID=10644;

RN SEQUENCE FROM N.A.

RA MEDLINE=82014878; PubMed=6269062;

RA Gardner R.C., Howarth A.J., Hahn P., Brown-Tuedi M., Shepherd R.J.,

RA Messing J.;

RT "The complete nucleotide sequence of an infectious clone of

RT Nucleic Acids Res. 9:2871-2888(1981).

RL Nucleic Acids Res. 9:2871-2888(1981).

CC -1- SIMILARITY: HIGH, WITH OTHER CAULIMOVIRUS ORF V.

CC -1- SIMILARITY: WITH RETROVIRAL POL/GAG POLYPROTEINS.

CC -1- SIMILARITY: THE PROTEIN BELONGS TO PEPTIDASE FAMILY A3.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC EMBL; V00140; -; NOT_ANNOTATED_CDS.

DR PIR; A04159; Q0CV5.

DR HSSP; P04585; 1KLM.

DR MEROPS; A03.001; -

DR InterPro; IPR000588; Peptidase_A3.

DR InterPro; IPR000477; RVTse.

DR Pfam; PF02160; Peptidase_A3; 1.

DR Pfam; PF00078; rvl; 1.

DR PRINTS; PR00731; CAULIMOPTASE.

KW Hydrolyase; Aspartyl protease; RNA-directed DNA polymerase;

KW Transferrase; Endonuclease; Polypeptide.

FT ACT_SITE 45 45

FT SIMILAR 40 130 TO RETROVIRUS GAG/POL. PROTEASE DOMAIN.

FT SIMILAR 260 620 TO RETROVIRUS GAG/POL. DNA POLYMERASE

SQ SEQUENCE 679 AA; 78669 MW; E21F8BC528C9DB8D CRC64;

Query Match 18.2%; Score 245.5; DB 1; Length 679;

Best Local Similarity 29.0%; Pred. No. 3e-15;

Matches 72; Conservative 49; Mismatches 90; Indels 37; Gaps 8;

QY 1 KEVVKLDEGIIYVAHSDWSPVSHVSKKGIYVVPNDKDELIPORITNG-YRMVDFR- 60

DB 264 KQIKELDLKYL-----KPSKS-PMAPAFLYNNEAKRKRGK-----RMVYVYKA 308

QY 61 LNKATRKDHYPPLFDIHLERLSKLTHRCFLDGYSSSQIPVAOSDEKTTFTCPFGTA 120

DB 309 MKKATGTADVYVLPNDKDELITLIRGKRTFSFCKSGFWQVLDLDESRPLTAFCTPOGAYE 368

QY 121 YRMPFGLCNAPATFORCMAIFSNFCENIVEVMDPSYVSSFDCLSLNDVLR 177

DB 369 MWVVPFGLKQAPSTFQRHDEAFERKFC-----CVYDDILVSNNEEDHLHYAMTLQ 424

QY 178 RCKDTNLVNGEKCHFWNNEGIVLGKHSERGIEVDRAKYDAIDKMPYPTDIDKIRSLGH 229

DB 425 KQNGHIIISKKAQ-----LFKKKTINFLGLDELTGTHRPGQHILHINKKPFDTLEDK 477

QY 230 KGIRSFGL 237

QY 230 KGI RSLG 237

ID	STANDARD	PR	0/9 AA
AC	P03554		
DT	21-JUL-1986 (Rel. 01, Created)		
DT	21-JUL-1986 (Rel. 01, Last sequence update)		
DT	20-AUG-2001 (Rel. 40, Last annotation update)		
DE	ENZYMATIC POLYPROTEIN [CONTAINS: ASPARTIC PROTEASE (EC 3.4.23.-)]		

GN Cauliflower mosaic virus (strain Strasbourg) (CaMV)
OS Viruses: Retroid viruses: Caulimovirus
OC

RP	RN	SEQUENCE FROM N.A.
		[1]

RA Franck A., Gullley H., Jonard G., Richards K., Hirth L.
RT "Nucleotide sequence of cauliflower mosaic virus DNA.";
RL Cell 21:285-294 (1980).

CC
CC
CC
CC
- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A3
- SIMILARITY: WITH REIKOVIRAL POL/GAG POLYPROTEINS.
CC

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/send_email_to_license@isb-sib.ch)

DR EMBL; V00141; CAA23460.1; -
DR PIR; A04159; QQC5.

```
DR      MEROPS; A03.001; -; Peptidase_A3
DR      InterPro; IPR000588;
DR      InterPro; IPR000477; RVTse
```

pfam; PF00078; rvt; 1.
PRINTS; PR00731; CAULIMOPTASE.
Hydrolyase; Acetartyl acetate; mu

Accession	Protein	Length	TO RETROVIRUS GAG/POL PROTEASE DOMAIN
KW	Transferase; Endonuclease; Polypotein.		
FT	ACT_SITE	45	
FT	SIMILAR	40	130

SEQUENCE 679 AA; 78629 MW; 9EE527BCD460B766 CRC64;

Best Local Similarity	29.08;	Pred. No. 3e-15;	Length 6/9;
Score	243.3;	DB 1;	
Match	10.20;		
Matrixes	72: Conservative	40: Mismatch	00: 1-4-11-32

```
QY      1 KEVVLDEGILYVAHSDMWSPVHSYKKGGITVPNDKDELPIQRITIGRMVIDFRK 60
```

QY 61 LNKATRKDHYELPFIDHMLERLSKLTNFCFLDGYSSFSQIPVAQSDQEXTTFTCPGTF 120

QY 121 YRRRPFGLCNAPATFORCM--MATEFNFCENIVEVEMDFSVYSSSPDDCLSNIDRVLO 177

QY 178 RCKDNLVLNGEKECHFMVNEGIVLGHKISERGEIVDKAK-----VDAIDKMPYP-TDI 229

```

OY      230 KGRSFG 237
        |::|||
Db      478 KOLORFLG 485

RESULT 11
RRPO_OENBE
ID      RRPO_OENBE          STANDARD;          PRT;          142 AA.
AC      P31843;
DT      01-JUL-1993 (Rel. 26, Created)
DT      01-JUL-1993 (Rel. 26, Last sequence update)
DT      15-DEC-1998 (Rel. 37, Last annotation update)
DE      RAN-DIRECTED DNA POLYMERASE HOMOLOG (REVERSE TRANSCRIPTASE HOMOLOG).
OS      Oenothera luteola (Bertero's evening primrose).
OC      Mitochondrion.
OC      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC      Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC      eurosids II; Myrtales; Onagraceae; Oenothera.
OX      NCBI_TaxID=3950;
RN      [1]
RA      SEQUENCE FROM N.A.
RA      Schuster W., Brennicke A.;
RT      "Plastid, nuclear and reverse transcriptase sequences in the
RT      mitochondrial genome of Oenothera: is genetic information transferred
RT      between organelles via RNA?";
RL      EMBO J. 6:2857-2863(1987).
CC      -I- SIMILARITY: TO REVERSE TRANSCRIPTASES.
-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
-----
DR      EMBL: X06034; CAZ9429.1; -.
DR      PIR: S01101; PROBHM.
DR      InterPro: IPR000477; RVISE.
DR      Pfam: PF00078; rvc; 1.
KW      Transactin; RNA-directed DNA polymerase; Mitochondrion.
SQ      SEQUENCE 142 AA; 16799 MW; 8593DF42Z6FA80C CRC64;

Query Match              18.0%; Score 242; DB 1; Length 142;
Best Local Similarity    37.9%; Pred. No. 1e-15;
Matches 50; Conservative 32; Mismatches 46; Indels 4; Gaps 2

OY      53 RMVIEFRKNNKTRKDHYLPFIDHMLERLSKLTHFCFLDGYSSFSQIPVAOSDOEKTF 112
        |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db      7 RMCIDVRALTYTKINKKYIIPRYDDLFDRLAQAQTWTKIDLNSGVQVRIAKODEBKTC 66

OY      113 TCDFGTFAVRMRPFGLCNAPATFFQRCMAAFESNFCNIYEVEMD---FSYVGSSPDDCL 169
        :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db      67 VTRYGSEFERVPVFGLTNALATPCNLMMNVLYEYLDHFVVYVLDDLVTYYISNSLHEHI 126

OY      170 SNMDFLRQCKD 181
        :|||::|:|:|
Db      127 KHL-RVRESKE 137

RESULT 12
POL_CAMVN
ID      POL_CAMVN          STANDARD;          PRT;          680 AA.
AC      Q00962;
DT      01-APR-1993 (Rel. 25, Created)
DT      01-APR-1993 (Rel. 25, Last sequence update)
DT      20-AUG-2001 (Rel. 40, Last annotation update)
DE      ENZYMATIC POLYPEPTIDE [CONTAINS: ASPARTIC PROTEINASE (EC 3.4.23.-)];
DE      ENDONUCLEASE; REVERSE TRANSCRIPTASE (EC 2.7.7.49)].
V.      Cauliflower mosaic virus (strain NW8153) (CAUV).
OS

```

CC	Viruses: Retroïd viruses; Caulimovirus.
OX	NCBI_TaxID=31557;
RN	[1]
RP	SEQUENCE FROM N.A.
RA	Chenault K.D., Steffens D.L., Melcher U.K.:
RT	"Nucleotide sequence of cauliflower mosaic virus isolate N8153."
RL	Plant Physiol. 100:542-545(1992).
CC	-1 SIMILARITY: HIGH, WITH OTHER CAULIMOVIRUS ORF V.
CC	-1 SIMILARITY: WITH RETROVIRAL POL/GAG POLYPROTEINS.
CC	-1 SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A3.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (see http://www.isb-slb.ch/announce/
CC	or send an email to license@isb-slb.ch).
CC	-----
DR	EMBL: M90541; AAA46358.1; -.
DR	HSSP: P04585; 1KLM.
DR	MEROPS: A03.001: -.
DR	InterPro: IPR000588; Peptidase.A3.
DR	InterPro: IPR000477; RVtse.
DR	Pfam: PF02160; Peptidase_A3; 1.
DR	Pfam: PF00078; rvt; 1.
DR	PRINTS: PR00731; CAULIMOPTASE.
KW	Hydrolase; Aspartyl protease; RNA-directed DNA polymerase;
KM	transferase; Endonuclease; Polyprotein.
FT	ACT_SITE 46 46
FT	BY SIMILARITY.
FT	SIMILAR 41 131
FT	TO RETROVIRUS GAG/POL PROTEASE DOMAIN.
FT	SIMILAR 261 621
FT	TO RETROVIRUS GAG/POL DNA POLYMERASE
FT	DOMAIN.
SEQUENCE	680 AA: 78665 MW; EC802E09647C221 CRC64;

```

Query Match      17.9%  Score 241.5;  DB 1;  Length 680;
Best Local Similarity 27.0%  Pred. No. 7.1e-15;
Matches 69;  Conservative 47;  Mismatches 87;  Indels 53;  Gaps 7;

QY      1 KEVAKLLDEGIIVHVAHSDWVSPVSHVSPKKGITVVPNDKDELIPORIT-----GY 52
      :::::|::|
Db      265 KQIKELLDLKV-----KPKSPHMAPVLVNNSEANGKNG 301

QY      53 RMVIDEFKLNKATRKDHYLPETIDHMLERLSKLHFCPLDGYSSFQSLIPAQSOEKTTF 112
      ||:::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      302 RMVAVYKAMKATVAGAYMLPKKDELTLTLIRKGIFFSFDCKSGFWOYLLDQESRPLTA 361

QY      113 TCPEPTATVYRMFPGCLNAPATFORCM--MAIISFNCENTVEVFMDSEVYSGSPDCL 169
      ||::|::|::|::|::|::|::|::|::|::|::|::|
Db      362 TCDPGHAEWVVVPEGLKQAPISIFORHMDAEFVEFKFC---CYVVDIYFVSNNEHDH 417

QY      170 SNDRVYORCKDNLVLNLSKCKHFVWNEGIVLGHKISERGIEVDKAK-----VDAIDK 222
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      418 LHVAMIIQCKNONGIILSKKKA-----LEKKINLGLGIDEGRHKPGHILEHINK 470

QY      223 MPYP-TDIKGISRFLG 237
      |::|::|::|
Db      471 PPDTLEDKQOLQRFLLG 486

RESULT 13
POL_CAMVD
ID POL_CAMVD STANDARD; PRT; 674 AA.
AC P03556;

DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ENZYMAATIC POLYPROTEIN [CONTAINS: ASPARTIC PROTEASE (EC 3.4.23.-);
DE ENONUCLEASE; REVERSE TRANSCRIPTASE (EC 2.7.7.49)].
GN V.
OS Cauliflower mosaic virus (strain D/H) (camv).
OC Viruses; Retroid viruses; Caulimovirus.

```


OS Figwort mosaic virus (strain Dxs) (FMV).
 OC Viruses: Retroid viruses; Caulimovirus.
 NCBI_TaxID=10650;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=83106468; PubMed=7152260;
 RA Balazs E., Guilley H., Jonard G., Richards K.;
 RT "Nucleotide sequence of DNA from an altered-virulence isolate D/H of
 the cauliflower mosaic virus.";
 RL Gene 19:239-249(1982).
 CC -1- SIMILARITY: HIGH, WITH OTHER CAULIMOVIUS ORF V.
 CC -1- SIMILARITY: WITH RETROVIRAL POL/GAG POLYPROTEINS.
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A3.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: M10376; AAA46350.1; ALT_INT.
 DR PIR: A04159; QOCV5.
 DR HSSP: P04585; 1KLM.
 DR MEROPS: A03.001; -
 DR InterPro: IPR000588; Peptidase_A3.
 DR InterPro: IPR000477; RVTse.
 DR Pfam: PF02160; Peptidase_A3; 1.
 DR Pfam: PF00078; rvt; 1.
 DR PRINTS: PR00731; CAULIMOPTASE.
 DR HydroLase: Aspartyl protease; RNA-directed DNA polymerase;
 KW Transferrase; Endonuclease; Polypeptide.
 FT ACT_SITE 40 130 TO RETROVIRUS GAG/POL PROTEASE DOMAIN.
 FT SIMILAR 260 620 TO RETROVIRUS GAG/POL DNA POLYMERASE
 FT
 SQ SEQUENCE 674 AA; 78164 MW; E004E3222D349E29 CRC64;
 Query Match 17.7%; Score 238.5; DB 1; Length 674;
 Best Local Similarity 28.6%; Pred. No. 1.3e-14;
 Matches 71; Conservative 49; Mismatches 91; Indels 37; Gaps 8;
 QY 1 KEVKLLDEGIIVNAHSDWSPVHSPKKGITVVPNDKDELIPQIIITGYRWIDFRK 60
 DB 259 KQIKELLDLGLT-----KPSKS-PMAPAPFLVNNNEKRRKK-----RMVNYKA 303
 QY 61 LNKATRKDHPLPFDIMHLERLSKLTJHFCFLDGYSSFSQIPVAOSDOEKTFTCPGTF 120
 DB 304 MNATVGDAYNPKNDELTLIRGKIFSSFDCKSGFWYVLDQESRPLAFTCPGCHYE 363
 QY 121 YRMPFGLCNAPATFORCMATFNSNCEINIEVFMDFSYVSSFDCLSLNDRVLQ 177
 DB 364 MNVFPGLKQAPSIQFQHMDEAFVRFKFC---CYVVDILVFSNNEEDHLHVMILQ 419
 QY 178 RCKDTLVINGEKCHPMVNEGIVLGHKISERGIEVDAK-----VDIDKMP-PTD 229
 DB 420 KCONHIIISKKKAQ-----LFFKKINFLGLEIDEGTHKPGCHILEHINKPPTLEDK 472
 QY 230 KGRISFLG 237
 DB 473 KQLOFLG 480
 RESULT 14
 POL_FMVND STANDARD: PRT; 666 AA.
 AC P09523;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-MAR-1989 (Rel. 10, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE ENZYMOLOGICAL PROTEIN (CONTAINS: ASPARTIC PROTEASE (EC 3.4.23.-);
 DE ENDOUCLEASE; REVERSE TRANSCRIPTASE (EC 2.7.7.49).
 GN V.

OS Figwort mosaic virus (strain Dxs) (FMV).
 OC Viruses: Retroid viruses; Caulimovirus.
 NCBI_TaxID=10650;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88040466; PubMed=3671088;
 RA Richins R.D., Scholthof H.B., Shephard R.J.;
 RT "Sequence of figwort mosaic virus DNA (caulimovirus group).";
 RL Nucleic Acids Res. 15:8451-8466(1987).
 CC -1- SIMILARITY: HIGH, WITH OTHER CAULIMOVIUS ORF V.
 CC -1- SIMILARITY: WITH RETROVIRAL POL/GAG POLYPROTEINS.
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A3.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: X06166; CAA29527.1; -
 DR PIR: S01283; S01283.
 DR HSSP: P04585; 1RT2.
 DR MEROPS: A03.001; -
 DR InterPro: IPR000588; Peptidase_A3.
 DR InterPro: IPR000477; RVTse.
 DR Pfam: PF02160; Peptidase_A3; 1.
 DR Pfam: PF00078; rvt; 1.
 DR PRINTS: PR00731; CAULIMOPTASE.
 DR HydroLase: Aspartyl protease; RNA-directed DNA polymerase;
 KW Transferrase; Endonuclease; Polypeptide.
 FT ACT_SITE 54 54
 FT
 SQ SEQUENCE 666 AA; 77081 MW; E65BC57D8FD2CA0F CRC64;
 Query Match 17.3%; Score 232.5; DB 1; Length 666;
 Best Local Similarity 28.6%; Pred. No. 4.8e-14;
 Matches 70; Conservative 52; Mismatches 92; Indels 31; Gaps 7;
 QY 1 KEVKLLDEGIIVNAHSDWSPVHSPKKGITVVPNDKDELIPQIIITGYRWIDFRK 60
 DB 257 KQIKELLDLGLT-PSKSQHMSPAF-----LVNEAEERRKK-----RMVNYKA 301
 QY 61 LNKATRKDHPLPFDIMHLERLSKLTJHFCFLDGYSSFSQIPVAOSDOEKTFTCPGTF 120
 DB 302 INQATIGDSHNLNMOELLTLIRGKIFSSFDCKSGFWYVLDQESOKLAFTCPOGHFQ 361
 QY 121 YRMPFGLCNAPATFORCMATFNSNCEINIEVFMDFSYVSSFDCLSLNDRVLQ 180
 DB 362 WKVFPGLKQAPSIQFQHMOTAL-NGADKFCMVYVDIIVFSNSELIDHNVHVAVLKIVE 420
 QY 181 DTNLVINGEKCHPMVNEGIVLGHKISERGIEVDAK-----VDIDKMP-yPTDIKI 222
 DB 421 KYGIIISKKKAN-----LFEKINFLGLEIDEGTHKPGCHILEHINKPPTLEDK 473
 QY 233 RSEFLG 237
 DB 474 QRFGLG 478
 RESULT 15
 POL_RTBVP STANDARD: PRT; 1675 AA.
 AC P27502; P27528;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE POLYPROTEIN (P194 PROTEIN) [CONTAINS: COAT PROTEIN; PROTEASE
 DE (EC 3.4.23.-); REVERSE TRANSCRIPTASE (EC 2.7.7.49); RIBONUCLEASE H
 DE (EC 3.1.26.4)].
 OS Rice tungro bacilliform virus (isolate Philippines) (RTBV).
 OC Viruses: Retroid viruses.

```

OX NCBI_TaxID=10655;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91252204; Pubmed=2041739;
RA Hay J.M., Jones M.C., Blakebrough M.L., Dasgupta I., Davies J.W.,
RA Hull R.;
RT "An analysis of the sequence of an infectious clone of rice tungro
RT bacilliform virus, a plant pararetrovirus.";
RL Nucleic Acids Res. 19:2615-2621(1991).
RN [2]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 503-526.
RX MEDLINE=92024093; Pubmed=1926781;
RA Ou R.D., Bhattacharya M., Laco G.S., de Kochko A., Rao B.L.S.,
RA Kaniewska M.B., Elmer J.S., Rochester D.E., Smith C.E.,
RA Beachy R.N.;
RT "Characterization of the genome of rice tungro bacilliform virus:
RT comparison with Comelina yellow mottle virus and caulimoviruses.";
RL Virology 185:354-364(1991).
CC -1- SIMILARITY: WITH COMELINA YELLOW MOTTLE VIRUS POLYPROTEIN, AND
CC WITH CAULIMOVIRUSES ORF V.
CC -1- SIMILARITY: WITH RETROVIRAL POL/GAG POLYPROTEINS.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY U33.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X57924; CAA40997.1; -.
CC EMBL: M65026; AAB03094.1; -.
CC PIR: C40785; C40785.
CC MEROPS: A03.002; -.
CC InterPro: IPR001995; Asp_prot_retrov.
CC InterPro: IPR001699; Asp_protease.
CC InterPro: IPR000477; RVase.
CC InterPro: IPR01878; ZnF_CCHC.
CC Pfam: PF000077; rvp; 1.
CC Pfam: PF000078; rvt; 1.
CC Pfam: PF000098; zf-CCHC; 1.
CC SMART: SM00343; Znf_C2HC; 1.
CC PROSITE: PS00141; ASF_PROTEASE; 1.
CC Polyprotein: RNA-directed DNA polymerase; Hydrolase;
CC Aspartyl protease; Transferase; Nucleotidyltransferase; Coat protein;
CC Endonuclease.
CC CHAIN 503 ? 33 KDA COAT PROTEIN.
CC FT ZN_FING 772 789 POTENTIAL.
CC FT ACT_SITE 987 987 PROTEASE (BY SIMILARITY).
CC FT CONFLICT 1292 1292 D -> E (IN REF. 2).
CC FT CONFLICT 1630 1630 Y -> H (IN REF. 2).
CC FT CONFLICT 1635 1635 S -> P (IN REF. 2).
CC SEQUENCE 1675 AA; 194081 MW; B24D7181463E466F CRC64;

Query Match 16.9%; Score 227.5; DB 1; Length 1675;
Best Local Similarity 26.0%; Pred. No. 4.2e-13;
Matches 68; Conservative 49; Mismatches 110; Indels 35; Gaps 6;

OY 1 KEVVKLLDEGIIVVAHSDWSPVHSVKKGGIT-----VVPNDKDLIPQRIITGY 52
DB 1201 KQIKELDKNLT-----KKADPTCRHRTAATFVRNHSEVAQKP----- 1239

OY 53 RNVIDFKLNKATRKDHYPLPFIIDHMLERLSKLTFCFLDGYSSFSOIPVAQSDQDKTTF 112
DB 1240 RIVYNNYRLNDNMHTDFNIPHKISMNLIQKANIIFSKFDLKAGFHHMKLKDKDFKDWTF 1299

OY 113 TCPEGFAFYRRMPFGCLNAPATFORCMATFSNFCENIVEMDPSVYSSPDDCLSNL 172
DB 1300 TCSEGLYTNVVCPIGANNACAFQRFQESFGDL--KFAITYIDDIILLASNNKEKIEHL 1357

OY 173 DRYLORCKDTNLVLNGEKHFVNEGIVLGHKISERGIEVDKAKVDAI---DKMPYPTDI 229

```

```

DB 1358 KIFENRVEVGCVLSSKSKMFLKEVEYLGVEIKRGKISLQPHIVDKIKKPKNKLT-L 1416
OY 230 KGIRSFGLHGGEYRPRIDFTK 251
DB 1417 KGLQAVLGLLNVARGVYIKDLSK 1438

```

Search completed: April 1, 2002, 22:47:31
Job time: 288 sec

Tue Apr 2 09:02:40 2002

us-09-586-106-63.rsp

Page 10


```

Db 1036 LNAASRKHFLPLFDQMLERLANHPYCLDGSGFQIPHPNDQKETTFTCPYGTFA 1095
QY 121 YRRMPFGLCNAPATFORCMMAIFSNFCENIVEFMDDFSVYSSFPDCLSLNDRYLQCK 180
Db 1096 YRRMPFGLCNAPATFORCMATISFDLIEWVEFMDDFSVYSSFPDCLSLNDRYLQCK 1155
QY 181 DTNVLNKGKCHFWNNEGIVGKHISERGIEVDKAKYDAIDKMPYPTDIKIGIRSLGHG 240
Db 1156 ETNVLNKGKCHFWNNEGIVGKHISERGIEVDKAKYDVMMQLOPKTKVDIRSFLGHAG 1215
QY 241 FYRRFIKDFPK 251
Db 1216 FYRRFIKDFPK 1226

RESULT 2
Q9SHM5 PRELIMINARY; PRT; 1862 AA.
AC Q9SHM5;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE F7F22.15.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
NCBI_Taxid=3702;
RN 111
RP SEQUENCE FROM N.A.
RA Chao Q., Shinn P., Brooks S., Buehler E., Dunn P., Khan S., Kim C.,
RA Walker M., Brooks S., Altafi H., Araujo R., Conn L., Conway A.B.,
RA Gonzalez A., Hansen N.F., Huizar L., Kremenetskaia I., Lenz C., Li J.,
RA Liu S., Lutos S., Rowley D., Schwartz J., Toriumi M., Vysotskaia V.,
RA Yu G., Davis R.W., Federspiel N.A., Theologis A., Ecker J.R.;
RA Genomic sequence for Arabidopsis thaliana BAC F7F22 from chromosome
RT I."
RT Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO RNA-DIRECTED DNA POLYMERASE (REVERSE
CC TRANSCRIPTASE).
DR EMBL: AC007534; AAF24529.1; -.
DR InterPro: IPR001584; RVE.
DR InterPro: IPR004477; RVTSE.
DR Pfam: PF00665; rve; 1.
DR Pfam: PF00078; rvt; 1.
DR RNA-directed DNA polymerase.
KW SEQUENCE 1862 AA; 212491 MW; 768944AA61EA519B CRC64;

```

```

Query Match 72.8%; Score 981; DB 10; Length 1862;
Best Local Similarity 71.3%; Pred. No. 5e-84;
Matches 179; Conservative 31; Mismatches 41; Indels 0; Gaps 0;

```

```

QY 1 KEVVLLDEGIIYHAHSDWSPVSHVPRKGGITVVPNDKDLIPQRIITGYRMYIDFK 60
Db 975 KEILLDLAGVYIPISDSWSPVSHVPRKGGITVVPNDKDLIPQRIITGYRMYIDFK 1034
QY 61 LNKATRKDHYPLPFDHMLERLSKLTTHFCFLDGSSFSQIPVAOSDOEKTTPCGPFA 120
Db 1035 LNAASRKHFLPLFDQMLERLANHPYCLDGSGFQIPHPNDQKETTFTCPYGTFA 1094
QY 121 YRRMPFGLCNAPATFORCMMAIFSNFCENIVEFMDDFSVYSSFPDCLSLNDRYLQCK 180
Db 1096 YRRMPFGLCNAPATFORCMATISFDLIEWVEFMDDFSVYSSFPDCLSLNDRYLQCK 1154
QY 181 DTNVLNKGKCHFWNNEGIVGKHISERGIEVDKAKYDAIDKMPYPTDIKIGIRSLGHG 240
Db 1156 ETNVLNKGKCHFWNNEGIVGKHISERGIEVDKAKYDVMMQLOPKTKVDIRSFLGHAG 1214
QY 241 FYRRFIKDFPK 251
Db 1215 FYRRFIKDFPK 1225

```

```

RESULT 3
Q9LPB1 PRELIMINARY; PRT; 1586 AA.
AC Q9LPB1;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE T32E20.9.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
NCBI_Taxid=3702;
RN 111
RP SEQUENCE FROM N.A.
RA Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C., Khan S.,
RA Kim C., Altafi H., Bel O., Chin C., Chou J., Choi E., Conn L.,
RA Conway A., Gonzalez A., Hansen N., Howing B., Koo T., Lam B.,
RA Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N., Nguyen M.,
RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,
RA Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,
RA Ecker J.R.;
RA Genomic sequence for Arabidopsis thaliana BAC T32E20 from chromosome
RT I."
RT Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO RNA-DIRECTED DNA POLYMERASE (REVERSE
CC TRANSCRIPTASE).
DR EMBL: AC020646; AAF79809.1; -.
DR InterPro: IPR001584; RVE.
DR InterPro: IPR004477; RVTSE.
DR Pfam: PF00665; rve; 1.
DR Pfam: PF00078; rvt; 1.
DR RNA-directed DNA polymerase.
KW SEQUENCE 1586 AA; 180875 MW; A9D3917EAFDAD790 CRC64;

```

```

Query Match 72.5%; Score 976; DB 10; Length 1586;
Best Local Similarity 71.3%; Pred. No. 1.2e-83;
Matches 179; Conservative 31; Mismatches 41; Indels 0; Gaps 0;

```

```

QY 1 KEVVLLDEGIIYHAHSDWSPVSHVPRKGGITVVPNDKDLIPQRIITGYRMYIDFK 60
Db 826 KEILLDLAGVYIPISDSWSPVSHVPRKGGITVVPNDKDLIPQRIITGYRMYIDFK 885
QY 61 LNKATRKDHYPLPFDHMLERLSKLTTHFCFLDGSSFSQIPVAOSDOEKTTPCGPFA 120
Db 886 LNAASRKHFLPLFDQMLERLANHPYCLDGSGFQIPHPNDQKETTFTCPYGTFA 945
QY 121 YRRMPFGLCNAPATFORCMMAIFSNFCENIVEFMDDFSVYSSFPDCLSLNDRYLQCK 180
Db 946 YRRMPFGLCNAPATFORCMATISFDLIEWVEFMDDFSVYSSFPDCLSLNDRYLQCK 1005
QY 181 DTNVLNKGKCHFWNNEGIVGKHISERGIEVDKAKYDAIDKMPYPTDIKIGIRSLGHG 240

```

Db 1006 ETNLVLMNEKHEWREGIVLGRKISEGIEVDKAKIDVNMOLQPPKTVKQIRSFGLHGAG 1065
 QY 241 FYRRFIKDFK 251
 Db 1066 FYRRFIKDFK 1076

RESULT 4
 Q9LHC0
 ID Q9LHC0 PRELIMINARY; PRT; 897 AA.
 AC Q9LHC0;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE RETROELEMENT POL. POLYPROTEIN-LIKE.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 RN NCBI_TaxID=3702;
 RP SEQUENCE FROM N.A.
 RA STRAIN=COLUMBIA;
 RA Nakamura Y.;
 RT "Structural Analysis of Arabidopsis thaliana Chromosome 3. III."; Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: TO RNA-DIRECTED DNA POLYMERASE (REVERSE TRANSCRIPTASE).
 DR EMBL: AP002054; BAB02630.1; -
 DR InterPro: IPR001584; Rve.
 DR InterPro: IPR00477; RVTse.
 DR Pfam: PF00665; rve; 1.
 DR Pfam: PF0078; rvt; 1.
 KW Polypeptide; RNA-directed DNA polymerase.
 SQ SEQUENCE 897 AA; 103002 MW; 007AGA04B2C7CE18 CRC64;

Query Match 70.3%; Score 947; DB 10; Length 897;
 Best Local Similarity 71.3%; Pred. No. 3.2e-81;
 Matches 174; Conservative 29; Mismatches 41; Indels 0; Gaps 0;

QY 8 DEGIYVHVASDWSVSPVHSVPKGGITVVPNDKDELIPORTITGYRMVIDEKLKATPK 67
 Db 66 DAGVIVPIDSSTWVSIVYCVPRKGGMTVANKEDILPRTITGHRMCIDYRKLNASK 125
 QY 68 DHPPLPEIDHMERLSEKLFHFCFLDGYSSFSQIPVQSDQEKTTFTCPGTFAYRRMPG 127
 Db 126 DHPPLPEIDHMERLSEKLFHFCFLDGYSSFSQIPVQSDQEKTTFTCPGTFAYRRMPG 185
 QY 128 ICAAPATFORCMAIFSNFCENIVEFMDPSVYSSFPDCLSLNDRVLQRCCKNTLVIN 187
 Db 186 ICAAPATFORCMAIFSNFCENIVEFMDPSVYSSFPDCLSLNDRVLQRCCKNTLVIN 245
 QY 188 GECCHWVNGEIVLGHKISRGIEVDKAKYDAIDKMPYPTDIKGISFSGHGFYRRFK 247
 Db 246 MECCHWVNGEIVLGHKISRGIEVDKAKYDAIDKMPYPTDIKGISFSGHGFYRRFK 305

QY 248 DFTK 251
 Db 306 DFTK 309

RESULT 5
 Q9SHM3
 ID Q9SHM3 PRELIMINARY; PRT; 1799 AA.
 AC Q9SHM3;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE F7F22.17.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 RN NCBI_TaxID=3702;
 RP SEQUENCE FROM N.A.
 RA Chao O., Shinn P., Brooks S., Buehler E., Dunn P., Khan S., Kim C., Walker M., Brooks S., Altafi H., Araujo R., Conn L., Conway A.B., Gonzalez A., Hansen N.F., Huizar L., Kremenetskaia I., Lenz C., Li J., Liu S., Luros S., Rowley D., Schwartz J., Toriumi M., Vysotskaia Y., Yu G., Davis R.W., Federspiel N.A., Theologis A., Ecker J.R.;
 RT "Genomic sequence for Arabidopsis thaliana BAC F7P22 from chromosome 1.";
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: TO RNA-DIRECTED DNA POLYMERASE (REVERSE TRANSCRIPTASE).
 DR EMBL: AC007534; ABF24531.1; -
 DR InterPro: IPR001584; Rve.
 DR InterPro: IPR00477; RVTse.
 DR Pfam: PF00665; rve; 1.
 DR Pfam: PF0078; rvt; 1.
 KW RNA-directed DNA polymerase.
 SQ SEQUENCE 1799 AA; 204493 MW; 0E5BACC0930579FC CRC64;

Query Match 70.2%; Score 946; DB 10; Length 1799;
 Best Local Similarity 69.7%; Pred. No. 9.9e-81;
 Matches 175; Conservative 31; Mismatches 39; Indels 6; Gaps 1;

QY 1 KEVKKLDEGIYVHVASDWSVSPVHSVPKGGITVVPNDKDELIPORTITGYRMVIDEKL 60
 Db 967 KEVKKLDEGIYVHVASDWSVSPVHSVPKGGITVVPNDKDELIPORTITGYRMVIDEKL 120
 QY 61 LNKATRKDHYPLPIDHMERLSEKLFHFCFLDGYSSFSQIPVQSDQEKTTFTCPGTF 120
 Db 1021 LNAASRKDHPPLPIDHMERLSEKLFHFCFLDGYSSFSQIPVQSDQEKTTFTCPGTF 1080
 QY 121 YRRMPFGLCNAPATFORCMAIFSNFCENIVEFMDPSVYSSFPDCLSLNDRVLQRC 180
 Db 1081 YRRMPFGLCNAPATFORCMAIFSNFCENIVEFMDPSVYSSFPDCLSLNDRVLQRC 1140
 QY 181 DTNLVLMNEKHEWREGIVLGRKISEGIEVDKAKIDVNMOLQPPKTVKQIRSFGLHG 240
 Db 1141 DTNLVLMNEKHEWREGIVLGRKISEGIEVDKAKIDVNMOLQPPKTVKQIRSFGLHG 1200
 QY 241 FYRRFIKDFK 251
 Db 1201 FYRRFIKDFK 1211

RESULT 6
 Q9MOT8
 ID Q9MOT8 PRELIMINARY; PRT; 724 AA.
 AC Q9MOT8;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE PUTATIVE ATHILA TRANSPOSON PROTEIN.
 GN ATG07660.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 RN NCBI_TaxID=3702;
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project.
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: TO RNA-DIRECTED DNA POLYMERASE (REVERSE TRANSCRIPTASE).
 DR EMBL: AL161506; CAB81136.1; -
 DR InterPro: IPR00477; RVTse.
 DR Pfam: PF0078; rvt; 1.
 KW RNA-directed DNA polymerase.
 SQ SEQUENCE 724 AA; 82390 MW; B80EB52DB35D5368 CRC64;

Query Match 62.4%; Score 841; DB 10; Length 724;

Best Local Similarity 66.8%; Pred. No. 2.8e-71;

Matches 155; Conservative 32; Mismatches 45; Indels 0; Gaps 0;

QY 1 KEVYKLLDEGIIYHVAHSDWSPVHSPKGGITVVPNDKDELIPORITGYRMVIDFRK 60
 DB 367 KEILKLDAGIYIPISDSTWSPVHSPKGGITVVPNDKDELIPORITGYRMVIDFRK 426
 QY 61 LNKATRKDHYPPLPTIDHMLERLSKLTFCFLDGYSSFSQIPVAOSDQKTFCTCPGTA 120
 DB 427 LNAASRKDHYPPLPTIDHMLERLSKLTFCFLDGYSSFSQIPVAOSDQKTFCTCPGTA 486
 QY 121 YRRMPFGICNAPATFORCMAIFSNFCENIVEFMDDFSVYSSSFDCLSLNDRYLQCK 180
 DB 487 YRRMPFGICNAPATFORCMAIFSNFCENIVEFMDDFSVYSSSFDCLSLNDRYLQCK 546
 QY 181 DTNVLNAGEKCHFWNEGIVLGHKISERGIEVDKAKYDAIDKMPYPTDIKIRSF 232
 DB 547 ETNVLNMEKCYFMVKEGIVLGHKISEGIEVDKAKYDAIDKMPYPTDIKIRSF 598

RESULT 7

ID 092PTS PRELIMINARY; PRT; 841 AA.

AC 092PTS; 01-MAY-1999 (TREMBLrel. 10, Created)
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE PUTATIVE POL. POLYPROTEIN WITH REVERSE TRANSCRIPTASE DOMAIN (PFAM: PF00078).
 GN F9B22.15.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Lin X., Kaul S., Shea T.P., Fujii C.Y., Shen M., Vanaken S.E.,
 RA Barnstead M.E., Mason T.M., Bowman C.L., Ronging C.M., Benito M.,
 RA Carrera A.J., Cressy T.H., Buell C.R., Town C.D., Nierman W.C.,
 RA Frazer C.M., Venter J.C.;
 RT "Arabidopsis thaliana chromosome II BAC F9B22 genomic sequence."
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: TO RNA-DIRECTED DNA POLYMERASE (REVERSE TRANSCRIPTASE).
 CC EMBL: AC006528; AAD19780.1;
 DR InterPro: IPR000477; RVTse.
 DR Pfam: PF00078; rvt; 2.
 KW Polyprotein, RNA-directed DNA polymerase.
 SQ SEQUENCE 841 AA; 96367 MW; 788BFAF2477E3F0F CRC64;

Query Match 61.3%; Score 826; DB 10; Length 841;

Best Local Similarity 62.2%; Pred. No. 8.9e-70;

Matches 156; Conservative 30; Mismatches 37; Indels 28; Gaps 2;

QY 1 KEVYKLLDEGIIYHVAHSDWSPVHSPKGGITVVPNDKDELIPORITGYRMVIDFRK 60
 DB 392 KEIMKLDAGIYIPISDSTWSPVHSPKGGITVVPNDKDELIPORITGYRMVIDFRK 451
 QY 61 LNKATRKDHYPPLPTIDHMLERLSKLTFCFLDGYSSFSQIPVAOSDQKTFCTCPGTA 120
 DB 452 LNSATRKDNFPLSFIDOMLERLSNQYCCFLDGYLGFGIILHPPDOEKTFTTCPEGTA 511
 QY 121 YRRMPFGICNAPATFORCMAIFSNFCENIVEFMDDFSVYSSSFDCLSLNDRYLQCK 180
 DB 512 YRRMPFGICNAPATFORCMAIFSNFCENIVEFMDDFSVYSSSFDCLSLNDRYLQCK 556
 QY 181 DTNVLNAGEKCHFWNEGIVLGHKISERGIEVDKAKYDAIDKMPYPTDIKIRSF 240

DB 557 DKHLVLMMEKSHFWYADGIVLGHKISEKGEVDRAKIE-----IMRFLGHAG 603

QY 241 FYRRFIKDPFK 251
 DB 604 FYRRFIKDPFK 614

RESULT 8

ID 022103 PRELIMINARY; PRT; 407 AA.

AC 022103; 01-JAN-1998 (TREMBLrel. 05, Created)
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE REVERSE TRANSCRIPTASE-LIKE PROTEIN (FRAGMENT).
 OS Vicia faba (Broad bean).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids I; Fabales; Fabaceae; Papilionoideae; Vicia.
 OX NCBI_TaxID=3906;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Kinoshita T., Wada H., Masaki I., Shimazaki K.;
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: TO RNA-DIRECTED DNA POLYMERASE (REVERSE TRANSCRIPTASE).
 CC EMBL: AB007466; BAA22787.1;
 DR InterPro: IPR000477; RVTse.
 DR Pfam: PF00078; rvt; 1.
 KW RNA-directed DNA polymerase.
 FT NON_TPR 1
 SQ SEQUENCE 407 AA; 47607 MW; 010378486EF0748A CRC64;

Query Match 57.5%; Score 774; DB 10; Length 407;

Best Local Similarity 70.9%; Pred. No. 3e-65;

Matches 139; Conservative 29; Mismatches 28; Indels 0; Gaps 0;

QY 56 IDRKLKATRKDHYPPLPTIDHMLERLSKLTFCFLDGYSSFSQIPVAOSDQKTFCTCP 115
 DB 3 IDRKLKATRKDHYPPLPTIDHMLERLSKLTFCFLDGYSSFSQIPVAOSDQKTFCTCP 62
 QY 116 FGTFAYRRMPFGICNAPATFORCMAIFSNFCENIVEFMDDFSVYSSSFDCLSLNDRY 175
 DB 63 FGTFAYRRMPFGICNAPATFORCMAIFSNFCENIVEFMDDFSVYSSSFDCLSLNDRY 122
 QY 176 LORCKDTNVLNAGEKCHFWNEGIVLGHKISEGIEVDKAKYDAIDKMPYPTDIKIRSF 235
 DB 123 LERCOESNLILNMEKCHFWNEGIVLGHKISYKIGIEVDQAKIEVSLHPTNEKIRSF 182
 QY 236 LGHGFYRRFIKDPFK 251
 DB 183 LGHGFYRRFIKDPFK 198

RESULT 9

ID 09C803 PRELIMINARY; PRT; 884 AA.

AC 09C803; 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE POLYPROTEIN, PUTATIVE.
 GN T4121.10.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;

RX MEDLINE=21016719; PubMed=11330712;
 RA Theologis A., Ecker J.R., Palm C.J.,
 RA White O., Alonso J., Altieri H., Araujo R., Bowman C.L., Brooks S.Y.,
 RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
 RA Chung M.K., Conn L., Conway A.B., Conway A.R., Cressy T.H., Dewar K.,
 RA Dunn P., Etgu P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y.,
 RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Hultar L.,
 RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
 RA Kim C.J., Koo H.L., Kremensskaja I., Kutz C.D.B., Kwan A., Lam B.,
 RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
 RA Lin X., Liu S., Liu Z.A., Luros J.S., Maiti R., Marziani A.,
 RA Miltitscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
 RA Pai G., Peterson J., Pham P.K., Ritzko M., Rooney T., Rowley D.,
 RA Sakano H., Salerg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
 RA Sun H., Tallon L.J., Tambunga G., Tortumi M.J., Town C.D.,
 RA Uterback T., Van Aken S., Vaysberg M., Vysotskaja V.S., Walker M.,
 RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.,
 RT "Sequence and analysis of chromosome 1 of the plant *Arabidopsis*
 RT *thaliana*."
 RL Nature 408:815-820(2000).
 CC -i SIMILARITY: TO RNA-DIRECTED DNA POLYMERASE (REVERSE
 CC TRANSCRIPTASE)
 DR EMBL: AC022456; AAG52026.1; -.
 DR InterPro: IPR000477; RVISE.
 DR Pfam: PF00078; rvt; 1.
 KW RNA-directed DNA polymerase.
 QO SEQUENCE 884 AA; 100352 MW; 5E57573BADDF8F98 CRC64;

Query Match	Score	DB	Length
47.88;	644;	10;	884;

Best Local Similarity / 58.1%; Pred. NO. 1.7E-32;
Matches 125; Conservative 30; Mismatches 38; Indels 22; Gaps 4

```

Qy 37 PNDKDELPOKILINGRYNVIDRFLKLNATKRDYDPLPIEDHMERLSKTLNFFCLDQSS 96
Db 503 PULK-AVKKEIM-----KLNATKRNHHPLOIGIDDLERLSNKHKTYCYLDQISG 552S
Qy 97 FSOIPVAOSDDEKITTCTPCTPFTAFVRRMRPCLCANAPATFOFCMAAIFSNPCENIVEYEMD 156
Db 553 FFQIPIDHDDQEKMFCTPYCTPAVSRRMPCLCANAPALFEKCMASIFLDMENIEYFMD 612
Qy 157 DFSYSSGFDDCLSLMDRYLORCKDTNLYLNGEKCHEVANEIGIYLGKHSRGIEYVAK 216
Db 613 DFSYSSGFACLENLRVYLARCEKKNLYLWMEKCHEFVQEGIYLGHRVSGAGIEYVNAK 672
Qy 217 VDAIDKMPYPPDIKIRBSFLHGGYRRFIEIDFKR 251
Db 673 IEVM-----ISLQALDS-----VNLRFVAFDFSK 695

```

	RESULT	10
09SK60		
ID	09SK60	PRELIMINARY; PRT; 622 AA.
AC	09SK60:	
DT	01-MAY-2000 (TREMBLrel. 13, Created)	
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)	
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)	
DE	PUTATIVE RETROELEMENT POL POLYPROTEIN.	
GN	AT2610690.	
OS	Arabidopsis thaliana (Mouse-ear cress).	
OC	Eurycoryta: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:	
OC	Spermatophytes: Magnoliophyta: eudicotyledons: core eudicots: Rosidae;	
OC	eucosids II: Brassicales; Brassicaceae; Arabidopsis.	
OX	NCBI_TaxID=3702;	
RN	111	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=CV. COLUMBIA;	
RX	MEDLINE=20083487; PubMed=10617197;	
RA	Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,	
RA	Fuill C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,	
RA	Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,	
RA	Cronin L.A., Shen M., Yanken S.E., Umayam L., Tallon L.J., Gill J.E.,	
RA	Adams W.D., Carreira A.J., Cressy T.H., Goodman H.M., Somerville C.R.,	

RA Copenhagen G.P., Preuss D., Niernan J.C., White O., Eisen J.A.,
 RA Salzberg S.L., Frazer C.M., Venter J.C.:
 RT "Sequence and analysis of chromosome II of *Arabidopsis thaliana*,"
 RL Nature 402:761-768(1999).
 CC -1 SIMILARITY: TO RNA-DIRECTED DNA POLYMERASE (REVERSE
 CC TRANSCRIPTASE).
 CC EMBL, AC006570; AAF18651.1; -.
 DR InterPro: IPR001584; Rye.
 DR InterPro: IPR000477; RYISe.
 DR Pfam: PF00665; Iye; 1.
 DR Pfam: PF00078; rvt; 1.
 KW Polypeptide: RNA-directed DNA polymerase.
 SQ SEQUENCE 622 AA; 71714 MW; BF69A4F1052AB579 CRC64;

Query Match	40.8%;	Score 549;	DB 10;	Length 622;
Best Local Similarity	46.2%;	Pred. No. 1.1e-43;		
Matches 121;	Conservative 30;	Mismatches 55;	Indels 56;	Gaps 6;

[illegible]

```

RESULT 11
09LJ78      PRELIMINARY;          PRT;          565 AA.
ID 09LJ78
AC 09LJ78;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)
DE SIMILARITY TO UNKNOWN PROTEIN.
OS Arabidopsis thaliana (mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-COLUMBIA;
RA Kaneo T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-COLUMBIA;
RX PubMed=10907853;
RA Nakamura Y.;
RT "structural analysis of Arabidopsis thaliana chromosome 3. II.
RT Sequence features of the regions of 4,251,695 bp covered by ninety P1,
RT TAC and BAC clones.";
RL DNA Res. 7:217-221(2000).
CC - SIMILARITY: TO RNA-DIRECTED DNA POLYMERASE (REVERSE
CC TRANSCRIPTASE).
DR EMBL: AP000732; BAB01210.1; -
DR Interpro: IPR000477; RVTse.
DR Pfam: PF00078; rvt. 1.
KW RNA-directed DNA polymerase.

```


SQ SEQUENCE 565 AA; 65009 MW; 87797E91062B55AD CRC64;
 Query Match 35.0%; Score 472; DB 10; Length 565;
 Best Local Similarity 44.6%; Pred. No. 2e-36;
 Matches 95; Conservative 17; Mismatches 37; Indels 64; Gaps 1;
 QY 18 SDWSPVHSPVKKGGITVVPNDKDELLPQRIITGYRMVIDERKLKATRKDHYPLFDH 77
 DB 236 SDWSPVHSPVKKGGITVVPNDKDELLPQRIITGYRMVIDERKLKATRKDHYPLFDH 250
 QY 78 MLERLSKLTHFCFLDGYSSFSQIPVPAOSDOKETFTGCFGFAATRRMFGICNAPATFOR 137
 DB 251 -----FFQIPHPDDQKFTFTCLYGFIAHRSFGICNAPATFOR 291
 QY 138 CMAAIFSNFCENIVEFMDFSVYSSGSSFDCLSLMDLVLOCRKDTNVLNGEKCHFWME 197
 DB 292 YMSAIFTFMEDFTFVEMDFSIYGFSEACLENLCKVLARCEKKNLLNMEKHFVYHE 351
 QY 198 GIVLGHKISERGTEVDKAKDAIDKMPYPTDIK 230
 DB 352 GIVLGHKISERGTEVDKAKDAIDKMPYPTDIK 384
 RESULT 12
 Q9VZ16 PRELIMINARY; PRT; 1062 AA.
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)
 DE CG2485 PROTEIN.
 GN MIDLINE-JUMPERPOL OR CG2485 OR CG17399.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NCBI_TaxID=7227;
 RP SEQUENCE FROM N.A.
 RC STRAIN-BERKELEY;
 RA MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.C., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champagne M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Ballou J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Beeson K.Y., Basu A., Baxendale J., Bayraktarglu L., Beasley E.M.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
 RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos P., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferrar C., Ferreira S., Fleischmann W.,
 RA Flier C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris J.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck M.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
 RA Jaitai M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Liao X., Lele Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svaitkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Wolley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster,"
 RL Science 287:2185-2195(2000).
 CC -1- SIMILARITY: TO RNA-DIRECTED DNA POLYMERASE (REVERSE
 CC TRANSCRIPTASE).
 CC EMBL: AE003485; AF048015.1;
 DR FLYBASE: FBgn0040019; midline-jumper-pol.
 DR InterPro: IPR001584; Rve.
 DR InterPro: IPR000477; RVTse.
 DR Pfam: PF00665; rve. 1.
 DR Pfam: PF00078; rvt. 1.
 KW RNA-directed DNA polymerase.
 SQ SEQUENCE 1062 AA; 123369 MW; 7BBCD41E88BFE4B2 CRC64;
 Query Match 33.8%; Score 455; DB 5; Length 1062;
 Best Local Similarity 38.4%; Pred. No. 1.9e-34;
 Matches 96; Conservative 52; Mismatches 88; Indels 14; Gaps 3;
 QY 2 EYVLLDEGIIVHSHDWSVHSPVKKGGITVVPNDKDELLPQRIITGYRMVIDERKL 61
 DB 205 QINKLLEQDILRH-SHSPWSAPVFLVPRK-----LDASNKK-----WRLVYDPRQL 250
 QY 62 NKATRKHYPLPFDHMLERLSKLTHTFCFLDGYSSFSQIPVPAOSDOKETFTGCFGFAAT 121
 DB 251 NDKTKRHYPMPTNINELDLGRAQYFSAIDLASGYHQIEVEPRDSKTAFAVAGHFEF 310
 QY 122 RRMFGICNAPATFORCMAAIFSNFCENIVEFMDFSVYSSGSSFDCLSLMDLVLOCRK 181
 DB 311 IRMPFGICNAPATFORCMAAIFSNFCENIVEFMDFSVYSSGSSFDCLSLMDLVLOCRK 370
 QY 182 TNVLNGEKCHFWMEFMDFSVYSSGSSFDCLSLMDLVLOCRKDTNVLNGEKCHFWME 241
 DB 371 ANKLQPDKSEFLKELEYLGHIVTEKGVKPNKTIETIAFPMPTRKIKNSFLGLGY 430
 QY 242 YRFRIDFKR 251
 DB 431 YRFRIDFKR 440
 RESULT 13
 Q9T0B7 PRELIMINARY; PRT; 1240 AA.
 AC Q9T0B7;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)
 DE PUTATIVE REVERSE-TRANSCRIPTASE-LIKE PROTEIN.
 GN T49.40 OR ATAG10580.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eucosids II; Brassicales; Brassicaceae; Arabidopsids.
 NCBI_TaxID=3702;
 RP SEQUENCE FROM N.A.
 RA Bevan M., Wedler H., Wambutt R., Bancroft I., Mewes H.W.,
 RA Mayer K.F.X., Schueller C.;
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Wedler H., Wedler E., Wambutt R., Mewes H.W., Lemcke K., Mayer K.F.X.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.


```

Db 677 LNRVTVKNKKYPLPRIDELLDLGATCESKIDLTSGYHQIPIAEADVRRKTAFTRYGHFE 736
QY 121 YRRMPEGLCNAPATFQRCMAIFSNFCENIVEVFMDDPSVYGSSFFDDCLSNDRVLOQCK 180
Db 737 FVMPFGLTNAPAFMRIMNSVFOEFLDEFYIIFIDILVYKSPEDHVLRRVMEKLR 796
QY 181 DTNVLNGEKCHFVNNEGIVLGHKISERGIEVDKAKYDAIDKMPYPTDIKGI RSFLGHG 240
Db 797 EEKLEARLSKCSFWQRRKMGFLGHIVSVEGVSVDPEKIEAIRDWP RPTNATEIRSFGLAG 856
QY 241 FYRRFIKDF 249
Db 857 YRRFVKG 865

```

Search completed: April 1, 2002, 22:46:11
 Job time: 1963 sec